

**MÁSTER UNIVERSITARIO EN BIOINFORMÁTICA**



VNIVERSITAT  
E VALÈNCIA

**TRABAJO DE FIN DE MÁSTER**

Caracterización de las diferencias de sexo en  
enfermedades cardiovasculares desde una  
perspectiva ómica

**AUTOR:**

**JOSE ANTONIO SÁNCHEZ MILÁN**

**TUTORES:**

**MARÍA PASCUAL MORA**

**MARTA R. HIDALGO GARCÍA**

**FRANCISCO GARCÍA GARCÍA**

**SEPTIEMBRE, 2020**





VNIVERSITAT  
E VALÈNCIA



Escola Tècnica Superior  
d'Enginyeria **ETSE-UV**

## **MÁSTER UNIVERSITARIO EN BIOINFORMÁTICA**

### **TRABAJO DE FIN DE MÁSTER**

# **Caracterización de las diferencias de sexo en enfermedades cardiovasculares desde una perspectiva ómica**

**AUTOR:**

**JOSE ANTONIO SANCHEZ MILÁN**

**TUTORES:**

**MARÍA PASCUAL MORA**

**MARTA R. HIDALGO GARCÍA**

**FRANCISCO GARCÍA GARCÍA**

---

**TRIBUNAL:**

PRESIDENTE/A:

VOCAL 1:

VOCAL 2:

**FECHA DE DEFENSA:**

**CALIFICACIÓN:**



## Resumen

Según la Organización Mundial de la Salud (OMS), las enfermedades cardiovasculares (ECV) son la principal causa de muerte a nivel mundial. Además, las ECV son una de las principales áreas donde se han investigado diferencias en la epidemiología, manifestación clínica, patofisiología, tratamiento y resultados en ambos sexos.

El objetivo del siguiente estudio es el de la identificación de los mecanismos moleculares que causan estas diferencias entre hombres y mujeres, con el fin de entender mejor la biología de las ECV que permita mejorar la precisión en el diagnóstico y proporcionar una mayor efectividad en el tratamiento personalizado.

Para ello se llevó a cabo una búsqueda sistemática de estudios con datos transcriptómicos de pacientes con ECV en las bases de acceso público, siguiendo los estándares establecidos en la declaración PRISMA para la búsqueda sistemática y metaanálisis. A continuación, para cada uno de los estudios seleccionados se realizó un análisis de expresión diferencial, enriquecimiento funcional de grupo de genes y análisis de rutas de señalización para encontrar estos mecanismos moleculares alterados en ambos sexos con ECV. Finalmente se integraron los resultados de todos los estudios mediante un metaanálisis a nivel de gen y función, que identificó un gran número de funciones moleculares alteradas afectando de forma diferencial a hombres y mujeres

**Palabras clave:** transcriptómica, enfermedades cardiovasculares, caracterización funcional, metaanálisis

---

## Abstract

According to the World Health Organization (WHO), cardiovascular diseases (CVDs) are the leading cause of mortality at the global level. Moreover, CVDs are one of the main areas where sex differences have been studied, finding differences in epidemiology, clinical manifestation, pathophysiology, treatment and outcomes in both sexes.

The aim of this study is the identification of the molecular mechanisms that lead to these differences between men and women in order to understand better the biology of CVDs and to allow us to improve the precision of the diagnosis and the effectiveness and personalization of the treatments.

By a systematic search of studies with transcriptomic data of patients with CVDs in the public databases following the PRISMA standards for systematic review and meta-analyses, we performed a differential expression analysis, gene set enrichment analysis and pathways analysis to find the different altered molecular mechanisms in both sexes with CVDs. The meta-analysis of these individual results showed a large group of altered molecular functions affecting differentially in men and women.

**Keywords:** transcriptomics, cardiovascular diseases, functional profiling, meta-analysis.



# Índice general

|  |           |
|--|-----------|
| <b>1. Introducción</b>   | <b>1</b>  |
| <b>1.1. Enfermedades cardiovasculares</b>  | <b>1</b>  |
| 1.1.1. Epidemiología las enfermedades cardiovasculares y su diferencia por sexos | 1         |
| 1.1.2. Estenosis Aórtica   | 3         |
| 1.1.3. Enfermedad de las arterias coronarias                                     | 4         |
| <b>1.2. Datos ómicos y tecnologías de alto rendimiento</b>                       | <b>7</b>  |
| 1.2.1. Transcriptómica   | 8         |
| 1.2.2. Análisis transcriptómicos   | 8         |
| 1.2.3. Metaanálisis de datos ómicos  | 10        |
| <b>2. Hipótesis y objetivos</b>  | <b>13</b> |
| <b>3. Material y métodos</b>   | <b>15</b> |
| <b>3.1. Revisión sistemática y selección de estudios</b>                         | <b>15</b> |
| <b>3.2. Análisis individual de los estudios</b>                                  | <b>16</b> |
| 3.2.1. Procesamiento de los datos  | 17        |
| 3.2.2. Análisis exploratorio   | 17        |
| 3.2.3. Análisis de expresión diferencial   | 18        |
| 3.2.4. Análisis de enriquecimiento funcional                                     | 18        |
| 3.2.5. Análisis de rutas de señalización   | 19        |
| <b>3.3. Metaanálisis:</b>  | <b>19</b> |
| 3.3.1. Metaanálisis de Genes   | 20        |
| 3.3.2. Metaanálisis de funciones   | 22        |
| <b>4. Resultados</b>   | <b>23</b> |
| <b>4.1. Revisión Sistemática y selección de estudios</b>                         | <b>23</b> |
| <b>4.2. Análisis bioinformático de estudios individuales</b>                     | <b>25</b> |
| 4.2.1. Procesamiento de los datos  | 25        |
| 4.2.2. Expresión diferencial   | 40        |
| 4.2.3. Enriquecimiento funcional   | 42        |
| 4.2.4. Análisis de rutas se señalización   | 48        |
| <b>4.3. Metaanálisis</b>   | <b>50</b> |
| 4.3.1. Metaanálisis de genes   | 50        |
| 4.3.2. Metaanálisis de funciones   | 52        |
| <b>5. Discusión</b>  | <b>57</b> |
| <b>6. Conclusiones</b>   | <b>61</b> |
| <b>Bibliografía</b>  | <b>63</b> |
| <b>Anexo I: Tablas</b>   | <b>67</b> |
| <b>Anexo II: Figuras</b>   | <b>73</b> |

## Abreviaturas:

|       |  |
|-------|--|
| ACV   | Accidente cerebrovascular                            |
| ADN   | Ácido desoxirribonucleico                            |
| ARN   | Ácido ribonucleico                                   |
| BH    | BH Benjamini & Hochberg                              |
| BP    | <i>Biological Process</i>                            |
| CCAA  | <i>Canonical Circuit Activity Analysis</i>           |
| CC    | <i>Cellular Component</i>                            |
| DL    | DerSimonian & Laird                                  |
| DM    | Diabetes Mellitus                                    |
| ECV   | Enfermedades Cardiovasculares                        |
| FAIR  | <i>Findable, Accessible, Interoperable, Reusable</i> |
| FDR   | <i>False discovery rate</i>                          |
| FE    | Efectos fijos  |
| GO    | <i>Gene Ontology</i>                                 |
| GSEA  | <i>Gene Set Enrichment Analysis</i>                  |
| HE    | Hedges   |
| HPO   | <i>Human Phenotype Ontology</i>                      |
| HS    | Schmidt & Hunter                                     |
| KEGG  | <i>Kyoto Encyclopedia of Genes and Genomes</i>       |
| logFC | Logaritmo de <i>fold change</i>                      |
| LOR   | Logaritmo de los <i>odds ratio</i>                   |
| MA    | Metaanálisis   |
| MF    | <i>Molecular Function</i>                            |
| MPA   | Mechanistic Pathway Activity                         |
| NCBI  | <i>National Center for Biotechnology Information</i> |
| PCA   | Análisis de componentes principales                  |

# 1.Introducción

## 1.1. Enfermedades cardiovasculares

Las enfermedades cardiovasculares (ECV) son un grupo de desórdenes del corazón y de los vasos sanguíneos, entre los que se incluyen, cardiopatía coronaria, enfermedades cerebrovasculares, arteriopatías periféricas, cardiopatía reumática, cardiopatía congénita, trombosis venosas profundas y embolias pulmonares.

Los ataques al corazón y accidentes cerebrovasculares (ACV) suelen ser fenómenos agudos que se producen principalmente por obstrucciones que impiden que la sangre fluya hacia el corazón o cerebro. La causa más frecuente en la formación de depósitos de grasa en las paredes de los vasos sanguíneos. También pueden deberse a la formación de coágulos de sangre o por hemorragias en los vasos cerebrales. Los ataques cardíacos y ACV suelen presentarse por la combinación de factores de riesgo.

Las principales causas son una dieta malsana, la inactividad física, el consumo de tabaco y alcohol. Estos factores comportamentales pueden manifestarse en forma de hipertensión arterial, hiperglucemia, hiperlipidemia y sobrepeso y obesidad. Estos últimos factores pueden medirse en los centros de atención primaria y son indicativos de un aumento en el riesgo de padecer una ECV. Están promovidos principalmente por los cambios sociales, económicos y culturales, como por ejemplo la globalización, la urbanización y el envejecimiento de la población. Otras causas son la pobreza, el estrés y los factores hereditarios.

Evitar o reducir los factores de riesgo comportamentales reducen el riesgo de padecer ECV. También se pueden administrar tratamiento farmacológico para la diabetes, hipertensión o hiperlipidemia con el fin de reducir el riesgo cardiovascular y prevenir los ataques cardíacos y accidentes cerebrovasculares.

### 1.1.1. Epidemiología las enfermedades cardiovasculares y su diferencia por sexos

Según los datos de la OMS, las ECV son la principal causa de muerte en todo el mundo. En 2015 murieron por enfermedades cardiovasculares 17,7 millones de personas, que representan el 31% de todas las muertes registradas en el mundo. De estas muertes, 7,4 millones se debieron a cardiopatía coronaria y 6,7 millones a accidentes cerebrovasculares.

Más de  $\frac{3}{4}$  partes de las defunciones por enfermedades cardiovasculares se producen en los países de ingresos medio-bajos. De los 17 millones de muertes de personas menores

de 70 años atribuibles a enfermedades no transmisibles, un 82% se corresponden a los países de ingresos medio y bajos y un 37% se deben a las enfermedades cardiovasculares.

La mayoría de las ECV pueden prevenirse actuando sobre factores de riesgo comportamentales, como el consumo del tabaco, dietas malsanas y obesidad, la inactividad física o el consumo nocivo de alcohol<sup>1</sup>.

Para las personas con un alto riesgo cardiovascular que presentan uno o más factores de riesgo, como hipertensión arterial, diabetes, hiperlipidemia o alguna ECV ya confirmada, son fundamentales la detección precoz y el tratamiento temprano, por medio de servicios de orientación o la administración de fármacos, según corresponda.

Numerosos estudios demuestran que hay diferencias en la prevalencia, diagnóstico, manifestaciones clínicas y patofisiología de diversas enfermedades cardiovasculares<sup>2,3</sup>, como por ejemplo angina e infarto agudo de miocardio o arteriopatía periférica, en relación al sexo. Otra enfermedad en la que se han descrito diferencias y desigualdades en cuanto al género es la cardiopatía isquémica crónica, en la que se describen diferencias en la edad de aparición, su fisiopatología, comorbilidad con otros factores de riesgo, especificidad de pruebas diagnósticas, tratamientos y resultados, siendo el pronóstico más desfavorable para las mujeres y hombres de grupos sociales desfavorecidos<sup>4</sup>.

Uno de los factores que puede aumentar esta prevalencia, es la presencia de un cuadro de síntomas atípicos en mujeres que causan errores en el diagnóstico o un retraso en el mismo, implicando una mayor mortalidad en mujeres, generando importantes diferencias en los tiempos de supervivencia<sup>4</sup>. La esperanza de vida es mayor en las mujeres, por lo que estas enfermedades suelen aparecer en etapas más tardías<sup>2,4</sup>.

También se suelen administrar tratamientos diferentes. Se prescriben más antiagregantes, estatinas y vasodilatadores periféricos a varones que a mujeres. Sin embargo, se prescriben más anticoagulantes y diuréticos a las mujeres. Se realizan más intervenciones en los varones y un mayor control de la dislipemia, por tanto, en mujeres la prevención de las ECV es más deficitaria<sup>4</sup>. Es por ello que son necesarios los estudios para el seguimiento a largo plazo de la prevención de las ECV en las mujeres y equiparar el abordaje de estas enfermedades de gran importancia y prevalencia en la sociedad actual en ambos sexos.

Como factores de riesgo, en hombres destaca el tabaquismo, mientras que en mujeres hay más comorbilidad con diabetes tipo 2, hipertensión arterial sistémica, obesidad y sedentarismo<sup>1,3</sup>. Además, en mujeres hay factores de riesgo específicos como son complicaciones después del embarazo, partos prematuros o neonatos con bajo peso al nacer, menopausia y terapia hormonal de reemplazo, anticonceptivos orales, síndrome de ovario poliquístico, entre otros<sup>3</sup>.

Generalmente, las mujeres padecen cardiopatía isquémica crónica unos 10 años después de los hombres, aunque la supervivencia de mujeres mayores de 65 años es similar a los hombres de la misma edad. En hombres, se reporta que el infarto de miocardio es una de las causas principales de muerte en la edad media de vida y padecen más enfermedades coronarias obstructivas y una mayor incidencia de muerte súbita<sup>4</sup>.

Entre las diferencias anatómicas, fisiopatológicas y hormonales por sexo, destacan las siguientes:

En mujeres:

- Los estrógenos intervienen en la vasodilatación y mejoran la función endotelial a través del óxido nítrico.
- Las hormonas sexuales están implicadas en cascadas adaptativas de señalización del miocardio a presión y sobrecarga de volumen, que incluyen la señalización del calcio, síntesis de óxido nítrico y mecanismos profilácticos.
- Padecen con mayor frecuencia enfermedad de un solo vaso, disfunción de la microvasculatura con alteraciones de la microcirculación, vasoespasmo coronario y disección de las arterias coronarias.
- Algunas pruebas diagnósticas tienen menor especificidad, como por ejemplo la prueba de esfuerzo.
- Tienen mayores complicaciones después de una intervención coronaria invasiva.
- Padecen con mayor frecuencia insuficiencia cardíaca diastólica con función sistólica normal e hipertrofia concéntrica ventricular con la fracción de eyección mantenida.
- Después de una reparación de la válvula aórtica presentan regresión más rápida de la hipertrofia y de las dimensiones del ventrículo izquierdo.
- Presentan miocardiopatía de Tako-Tsubo asociada a situaciones de estrés importante.
- Están más expuestas a los fármacos, en la que pueden padecer más efectos adversos de los betabloqueantes, de los inhibidores de la enzima convertidora de angiotensina y los diuréticos.

En hombres:

- Presentan más estenosis del tronco coronario y enfermedad de múltiples vasos.
- Desarrollan con mayor frecuencia insuficiencia cardíaca con fracción de eyección reducida e hipertrofia excéntrica con ventrículo dilatado.
- Interpretan más rápido sus síntomas como secundarios a la insuficiencia cardíaca y por tanto inician el tratamiento más temprano y de forma más específica.

A continuación, se describen algunas de las enfermedades cardiovasculares más importantes en las que se han descrito diferencias por sexo y género:

### **1.1.2. Estenosis Aórtica**

La estenosis aórtica es una de las enfermedades valvulares más frecuente y grave<sup>5</sup>. En la estenosis aórtica, se produce una estrechez en la válvula que restringe el flujo de sangre desde el ventrículo izquierdo a la aorta y puede producir aumento de presión en la aurícula izquierda.

En la población Europea, la prevalencia de estenosis aórtica es del 5% entre personas de 70 a 79 años y hasta el 10% en personas mayores de 80 años<sup>2</sup>. En el caso de estenosis aórtica debido a un defecto congénito de válvula bicúspide es hasta 3 veces más frecuente

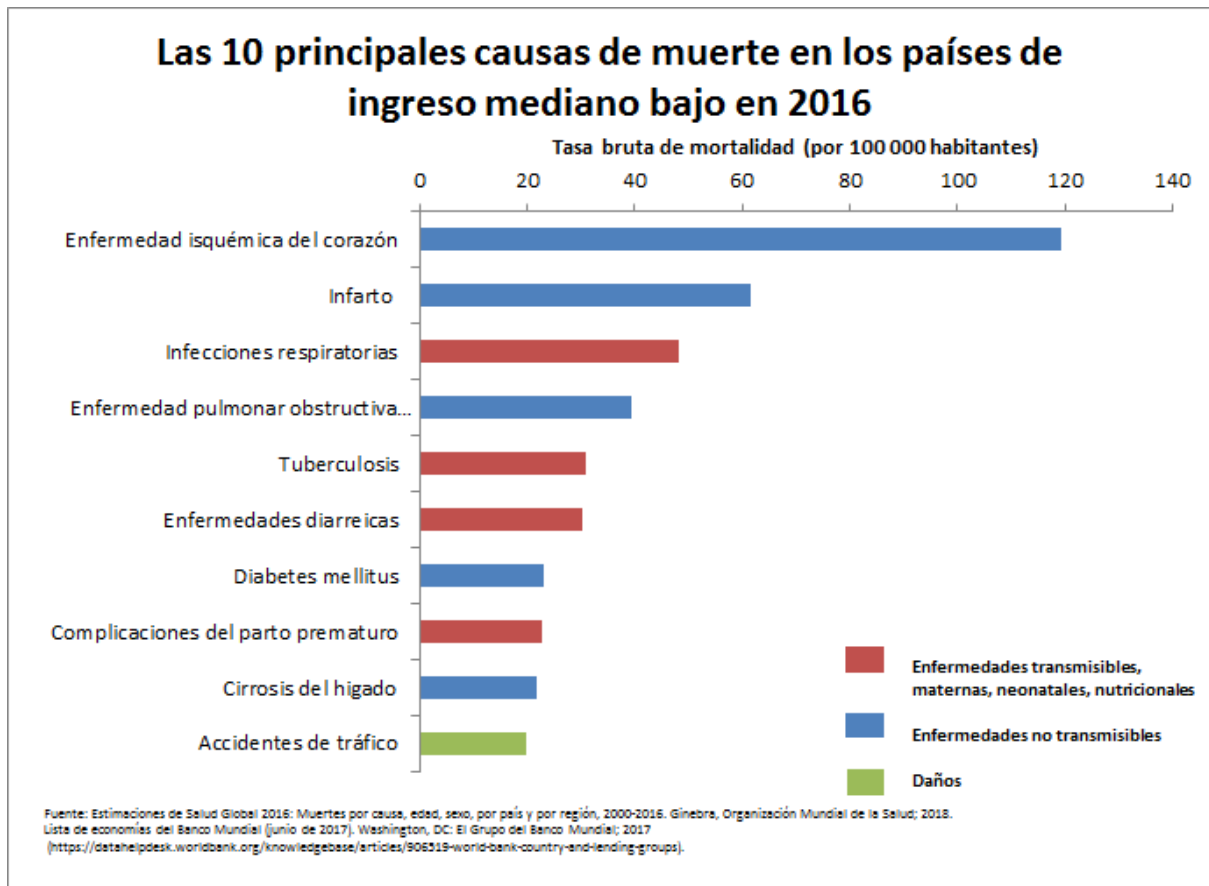
en hombres que en mujeres, mientras que la estenosis aórtica degenerativa es más frecuente en mujeres que en hombres.

A pesar de que muchos casos se deben por un defecto congénito llamado válvula aórtica bicúspide, la mayoría de casos se desarrolla durante el envejecimiento por el efecto del calcio. A continuación se describen los aspectos fundamentales de esta patología:

- Síntomas. No siempre se producen síntomas, pero entre ellos destaca falta de aliento, dolor en el pecho (angina), debilitamiento (síncope), palpitaciones, reducción en la capacidad de realizar actividades que requieren un esfuerzo leve y soplo cardíaco<sup>6</sup>.
- Factores de riesgo. La estenosis aórtica afecta principalmente a gente de edad avanzada como resultado de la cicatrización y acumulación de calcio en la cúspide de la válvula. La causa más frecuente en individuos jóvenes es por defecto congénito (válvula bicúspide) o por un crecimiento irregular en la que la abertura no crece junto al corazón, por lo que el corazón tiene que trabajar más. Con los años, la válvula defectuosa se vuelve rígida y estrecha debido a la acumulación de calcio.
- Diferencias en cuanto al sexo. La manifestación clínica no suele diferir entre hombres y mujeres, sin embargo, como consecuencia del aumento de la presión de recarga en el ventrículo izquierdo, las mujeres suelen desarrollar una hipertrofia concéntrica con una cavidad interna menor y una pared muscular más gruesa que en hombres<sup>2</sup>.
- Diagnóstico. El diagnóstico de estenosis aórtica no difiere entre hombres y mujeres, aunque la medida de la calcificación medida por tomografía es más marcada en hombres que en mujeres.
- Tratamiento. Se ha reconocido mayores complicaciones y fallecimientos en mujeres que han recibido tratamiento quirúrgico de reemplazo de válvula aórtica<sup>4</sup>. Sin embargo, tratamientos como la implantación de un transcáteter de válvula aórtica ha reportado mejoras en cuanto a la supervivencia<sup>2</sup>.

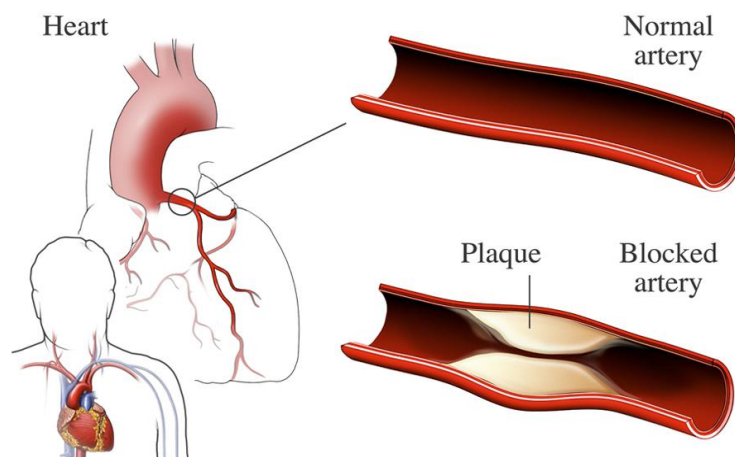
### 1.1.3. Enfermedad de las arterias coronarias

Según la OMS, las principales causas de muerte a nivel mundial son la cardiopatía isquémica y el accidente cerebrovascular, ocasionando 15,2 millones de defunciones en 2016, manteniéndose también como principales causas de mortalidad durante los últimos 15 años.



**Figura 1.1:** Principales causas de muerte en países con ingreso medio según datos de la OMS. Fuente: <https://www.who.int/es/news-room/fact-sheets/detail/the-top-10-causes-of-death>

Esta enfermedad isquémica está ocasionada principalmente por la enfermedad arterial coronaria, en la que se forman placas en las venas coronarias donde se acumulan lípidos y células inflamatorias y con ello, disminuye su capacidad de aporte de oxígeno y nutrientes al músculo cardíaco, produciendo así un daño y fibrosis del músculo cardíaco y también una disminución del aporte sanguíneo<sup>7</sup>.



**Figura 1.2:** Formación de una placa de ateroma en la enfermedad de las arterias coronarias. Fuente: <https://www.nhlbi.nih.gov/health-topics/ischemic-heart-disease>

Entre los factores que desencadenan la enfermedad, destacan los factores modificables como son la Diabetes Mellitus, hipertensión, consumo de tabaco, hiperlipidemia, obesidad y estilo de vida sedentario, mientras que entre los factores no modificables destacan el sexo, la edad y la predisposición genética<sup>7,8</sup>.

#### Diferencias en cuanto al sexo:

También se han reportado diferencias en cuanto al sexo<sup>2,9</sup>. En cuanto a su manifestación, suele desarrollarse de 7 a 10 años más tarde en mujeres con respecto a los hombres<sup>9</sup>. Generalmente, se manifiesta como síndrome agudo coronario, infarto de miocardio con elevación del segmento ST o sin el elevamiento del segmento ST entre 3 y 4 veces más en hombres que en mujeres por debajo de los 60 años. Sin embargo, por encima de los 75 años, la mayoría de casos se reportan en mujeres.

Como se ha comentado anteriormente, la enfermedad arterial isquémica del corazón afecta tanto a hombres como a mujeres y también se han reportado diferencias clínicas en cuanto al sexo. Los hombres suelen presentar más casos de enfermedad arterial coronaria del tipo obstructivo, mientras que las mujeres suelen presentar más casos de enfermedad arterial del tipo no obstructivo y debido principalmente a alteraciones vasculares, como disfunción endotelial o espasmos<sup>3</sup>.

Además del sexo, la edad también es un factor importante para la enfermedad isquémica del corazón. En los hombres, el riesgo de padecer la enfermedad isquémica del corazón aumenta a partir de los 45 años. En las mujeres, antes de la menopausia, presentan un menor riesgo de padecer la enfermedad que los hombres. Sin embargo, a partir de los 55 años, el riesgo de padecer la enfermedad se equipara al de los hombres. Esto es debido principalmente a que disminuye el efecto protector del estrógeno después de la menopausia<sup>8,9</sup>.

En la mujer, la diabetes es el factor de riesgo que mayor afecta al desarrollo de enfermedad arterial coronaria, en relación incluso a los hombres. Además, las mujeres pueden presentar factores de riesgo específicos, como por ejemplo diabetes gestacional que también aumenta el riesgo de padecer la enfermedad<sup>2</sup>.

#### Diagnóstico:

También se ha reportado un diagnóstico diferencial entre hombres y mujeres. Generalmente, las mujeres acuden más tarde a los centros de salud, por lo que puede conllevar a un diagnóstico más tardío y a un mayor progreso de la enfermedad<sup>4</sup>. Se han descrito una serie de marcadores que pueden contribuir a un mejor diagnóstico diferencial entre hombres y mujeres, como es el caso de la medición de los niveles de troponina I en el caso de infarto agudo de miocardio<sup>2</sup>.

Además, la interpretación de los resultados no invasivos es menos fiable en mujeres que en hombres ya que suelen ser resultados anormales<sup>4</sup>. Por ejemplo, en los

electrocardiogramas tienen menor sensibilidad y especificidad para pequeñas pruebas de esfuerzo en mujeres, ya que la puntuación ha sido ajustada principalmente a hombres.

La angiografía tampoco es fiable, ya que las mujeres por debajo de los 60 años no suelen presentar alteraciones obstructivas como ocurre en los hombres y por tanto, suelen ser infra diagnosticadas.

## 1.2. Datos ómicos y tecnologías de alto rendimiento

El ADN o ácido desoxirribonucleico es la molécula que tiene la función fundamental de almacenamiento de la información genética para construir otros componentes fundamentales de las células, como son proteínas y moléculas de ARN o ácido ribonucleico, fundamentales para el desarrollo y funcionamiento de los organismos vivos, además de ser responsable también de la transmisión hereditaria<sup>10</sup>.

Este ADN está formado por 2 hélices antiparalelas compuestas por nucleótidos. Cada nucleótido está compuesto por una base nitrogenada, que son la Adenina (A), Timina (T), Citosina (C) y Guanina (G), además de presentar una desoxirribosa y un grupo fosfato. La disposición de estas bases nitrogenadas a lo largo de la cadena de ADN es la que codifica la información genética.

Para que esta información genética pueda ser utilizada para realizar las funciones de la célula, es necesario que se copie la información de ADN a ARN o ácido ribonucleico, proceso llamado transcripción. Este ARN sirve a su vez de molde para la síntesis de proteínas a partir del proceso de traducción o bien realizar un gran número de funciones reguladoras<sup>10</sup>.

El proceso desde la secuencia que contiene la información genética (gen) hasta pasar por molécula de ARN mensajero en el proceso de transcripción y finalmente a proteína en el proceso de traducción, es el dogma central de la biología molecular<sup>11</sup>.

Con el fin de estudiar cada uno de estos procesos a fin de entender las funciones moleculares, se han desarrollado diferentes ramas en biología molecular, centradas en cada proceso. De esta forma, la genómica es la ciencia que estudia el genoma o conjunto de genes de un organismo. La transcriptómica es la ciencia que se encarga del estudio del conjunto de ARN del organismo y la proteómica la ciencia encargada del estudio de las proteínas del organismo.

Este conjunto de ciencias reciben el nombre de ómicas y puesto que se centran en el estudio de todo el conjunto de moléculas de cada apartado, se genera una gran cantidad de información que es necesaria de almacenar en bases de datos.

Debido al incremento de usos de tecnologías de alto rendimiento, como son la secuenciación masiva y los *microarrays*, se ha generado una gran cantidad de información cuyo procesamiento y análisis permiten identificar y comprender los procesos que ocurren en el organismo y cuya alteración puede ser causante de las enfermedades<sup>12</sup>.

### 1.2.1. Transcriptómica

Como se ha comentado en el apartado anterior, la transcriptómica es la ciencia encargada del estudio del conjunto de ARN de una célula o grupo de células en un momento dado. Este ARN mensajero procedente de la transcripción de ADN puede contener la información para codificar una proteína y por tanto, es ARN codificante. También puede tener una función reguladora y no dar lugar a proteínas y ser por tanto ARN no codificante.

La expresión de ARN mensajero varía con las condiciones del medio frente a respuesta a estímulos externos y gracias a su estudio, se permite aumentar el conocimiento de los mecanismos alterados en ciertas enfermedades, en la que las funciones celulares normales se ven alteradas.

Hay 2 formas principales de estudiar la expresión genética y son mediante el uso de *microarrays* y las tecnologías de secuenciación de ARN.

Los *microarrays* son pequeños chips donde hay insertadas secuencias de oligonucleótidos<sup>13</sup>. Normalmente las muestras de ARN a estudiar son marcadas con fluorescencia y dependiendo de la secuencia, hibridan con estos oligonucleótidos. La intensidad de fluorescencia detectada es proporcional a la expresión de ese gen o secuencia de ARN mensajero.

La secuenciación de ARN o RNA-seq es la técnica que permite detectar todos los ARN mensajeros presentes en ese momento en la célula o grupo de células<sup>14</sup>. Permite identificar tanto genes conocidos como desconocidos. Previamente a la secuenciación, la molécula monocatenaria de ARN mensajero es transformada a ADN complementario (ADNc) mediante transcripción inversa o retrotranscripción. Seguidamente se amplifican estas copias y se alinean con el genoma de referencia, con el fin de determinar los niveles de expresión de cada gen.

### 1.2.2. Análisis transcriptómicos

Hay diferentes estrategias de análisis de datos ómicos y se deben adaptar al tipo de dato a analizar, así como también al objetivo del estudio.

En concreto, con los datos transcriptómicos, se requiere un preprocesamiento de los datos, que dependerá de la tecnología de alto rendimiento empleada. Como se ha comentado anteriormente, los datos obtenidos de estudios transcriptómicos provienen de tecnología de

microarrays y de RNA-seq, principalmente<sup>13,14</sup>. De estos estudios, se obtienen matrices que contienen la cuantificación de la unidad biológica de interés. Las columnas contienen la información de las muestras y las filas, las sondas empleadas o genes.

En el caso de *microarrays*, contienen la información del nivel de transcripción de cada sonda, mientras que en el caso de RNA-seq, contienen la información de los conteos de los transcritos.

Las estrategias de análisis de datos transcriptómicos<sup>15</sup> más comunes son:

- 1) **Análisis de expresión diferencial:** su objetivo es la identificación de genes o transcritos que varían su expresión entre los diferentes grupos experimentales.
- 2) **Predicción de clases:** a partir de un conjunto de datos que sirven de entrenamiento para el modelo, con perfiles de transcripción diferentes entre las diferentes condiciones, es posible determinar un predictor supervisado que clasifique nuevas muestras.
- 3) **Análisis de clúster:** su objetivo es el de agrupar las diferentes muestras o grupos de genes, a partir de un patrón de expresión común. Este tipo de análisis también se denomina clasificación no supervisada.

### **Caracterización funcional:**

A partir de la información obtenida en los análisis comentados anteriormente, es posible identificar genes o transcritos que pueden tener un patrón de expresión similar y determinar de esta forma, hipótesis sobre los mecanismos que pueden estar siendo alterados en las condiciones de enfermedad. Por lo tanto, es interesante además de determinar qué genes o transcritos están alterados ante estas situaciones, detectar las funciones moleculares en las que participan estos genes, mediante diferentes estrategias de caracterización funcional.

Para ello, se requiere la información disponible en bases de datos, en la que se describen las funciones o procesos biológicos en los que están implicados los genes, haciendo de esta forma más fácilmente interpretables los resultados. Algunas de las bases de datos más utilizadas para llevar a cabo la caracterización funcional son las siguientes:

- *Gene Ontology* (GO)<sup>16</sup>: Contiene la información que relaciona los genes con funciones específicas en las que intervienen. Los términos están organizados jerárquicamente en 3 ontologías que describen procesos biológicos (BP), componentes celulares (CC) y funciones moleculares (MF).
- *Kyoto Encyclopedia of Genes and Genomes* (KEGG)<sup>17</sup>: Uno de los aspectos más interesantes de esta base de datos es que dispone de información acerca de las rutas metabólicas en las que participan los genes, organizados en forma de red.
- *Reactome*<sup>18</sup>: Al igual que la base de datos de KEGG, dispone de información sobre las rutas metabólicas en las que participan los genes, además de poseer herramientas de análisis bioinformáticas de estas rutas.

- *Human Phenotype Ontology* (HPO)<sup>19</sup>: Es una base de datos que contiene información anotada de alteraciones fenotípicas asociada a enfermedades, pudiendo también relacionar los genes con estas alteraciones.

De esta forma, combinando la información disponible de estas bases de datos junto a los resultados obtenidos del análisis de expresión diferencial, es posible llevar a cabo la caracterización funcional a través de diferentes métodos.

Por un lado, están los análisis de enriquecimiento o sobrerrepresentación<sup>20</sup>, en los que se caracterizan las funciones de un grupo específico de genes (por ejemplo: genes sobreexpresados en un grupo enfermo respecto un grupo control). En este análisis, se comparan las funciones anotadas de este grupo de genes con las funciones de otro grupo de genes de referencia (por ejemplo el genoma completo humano o una selección de genes de interés). De esta forma se evalúa estadísticamente la proporción de genes asociados a una determinada función, pudiendo detectarse sobrerrepresentación de genes asociados a dicha función. Tiene la limitación de que se estudia la lista específica de genes procedente de una selección realizada a partir de un punto de corte.

Otra aproximación es el análisis de grupo de genes, como por ejemplo el GSEA<sup>21</sup> (*Gene Set Enrichment Analysis*). En esta metodología se utiliza la lista completa de genes del estudio y se ordenan según su nivel de expresión diferencial. De esta forma se determina si los genes que participan en una misma función tienden a agruparse y a presentar patrones de expresión alta (sobrerexpresados) o baja (infraexpresados) en los diferentes grupos a estudiar. Al final se obtiene la lista de funciones significativas que están siendo alteradas.

Siguiendo con los métodos descritos anteriormente surgen las estrategias de análisis de actividad de rutas metabólicas, como el *Canonical Circuit Activity Analysis* (CCAA)<sup>22</sup>. Este método tiene en cuenta los genes que participan en una misma ruta metabólica y su regulación. De esta forma, se puede obtener un valor de actividad de la ruta y comprobar de esta forma si las alteraciones producen una mayor actividad en la ruta o una disminución, entre las diferentes condiciones.

### 1.2.3. Metaanálisis de datos ómicos

El metaanálisis (MA) surge como una herramienta para combinar los resultados de diferentes estudios para obtener así una mayor evidencia científica<sup>23</sup>. Esta metodología proporciona una mayor potencia estadística, obteniendo una medida combinada del efecto de interés. Su aplicación mejora la precisión en la estimación del efecto y reduce la probabilidad de falsos negativos<sup>24</sup>.

Un paso previo al MA es la revisión sistemática o la síntesis de estudios relacionados pero independientes, sobre un área de conocimiento, como una determinada enfermedad o un grupo de patologías<sup>25</sup>. Algunos de estos estudios independientes pueden llegar a tener

conclusiones contradictorias y por tanto el metaanálisis puede ser utilizado para estimar el efecto promedio o determinar el grupo de estudios que apuntan a la misma dirección, con el fin de obtener información valiosa en términos de evidencia científica <sup>23</sup>.

Inicialmente fue aplicado en ciencias sociales y psicología, pero a partir de los años 80 aumentó su uso creciente en ciencias médicas, y en los 90 empezaron a crecer las publicaciones de revisión y metaanálisis<sup>26</sup>.

El abaratamiento de las técnicas de alto rendimiento como los *microarrays* y la secuenciación de nueva generación ha propiciado un gran aumento del uso de estas técnicas en el campo de la biomedicina<sup>24</sup>, y con ello se han generado una gran cantidad de datos que han sido depositados en bases de datos de acceso público en las dos últimas décadas. Por este motivo, el metaanálisis es una potente herramienta para el análisis de estos datos, con el fin de detectar biomarcadores asociados a enfermedades que proporcionen una mejora de los diagnósticos y un incremento de la precisión en los tratamientos.



## 2. Hipótesis y objetivos

Las enfermedades cardiovasculares tienen un alto impacto de salud en la población general. Además en algunas de ellas se presentan diferencias específicas por sexo, que se han descrito a nivel epidemiológico, así como en el grado de severidad y en el desarrollo de este grupo de patologías. Nuestra hipótesis de estudio es que existen mecanismos diferenciales por sexo en varias de estas enfermedades cardiovasculares. Su identificación y caracterización permitiría ser más precisos en los procesos diagnósticos y de tratamiento para hombres y mujeres.

El objetivo principal del siguiente estudio es la identificación de los mecanismos moleculares alterados en estas enfermedades, que afectan de forma diferencial a hombres y mujeres, mediante un abordaje *in silico*. Concretamente, nos centraremos en las enfermedades: estenosis aórtica y la enfermedad de las arterias coronarias. Para ello, se han propuesto los siguientes objetivos específicos:

- 1) Revisión sistemática y selección de estudios de enfermedades cardiovasculares en repositorios de acceso público GEO y ArrayExpress para la obtención de datos de *microarrays* y RNA-Seq.
- 2) Detección y evaluación de los factores de riesgo que pueden intervenir en cada enfermedad.
- 3) Análisis bioinformático individual para los datos de cada estudio seleccionado, incluyendo análisis exploratorio, análisis de expresión diferencial, enriquecimiento funcional y análisis de rutas.
- 4) Identificación de genes en el conjunto de los estudios seleccionados que presenten un perfil diferencial por sexo.
- 5) Caracterización funcional de las diferencias por sexo en las enfermedades evaluadas, mediante el metaanálisis de funciones biológicas, rutas de señalización y ontologías fenotípicas.



## 3. Material y métodos

En este estudio *in silico* se han utilizado datos transcriptómicos procedentes de bases de datos de acceso público. Nos centraremos en la integración de los resultados de todos los estudios tanto a nivel de gen como de función. Durante las distintas fases de la estrategia de análisis, se ha utilizado el lenguaje de programación R, generando diversos scripts que están disponibles en [gitlab](#). Además se anexa un listado de las librerías utilizadas en la realización de cada proceso (Anexo I, tabla a1 y a2).

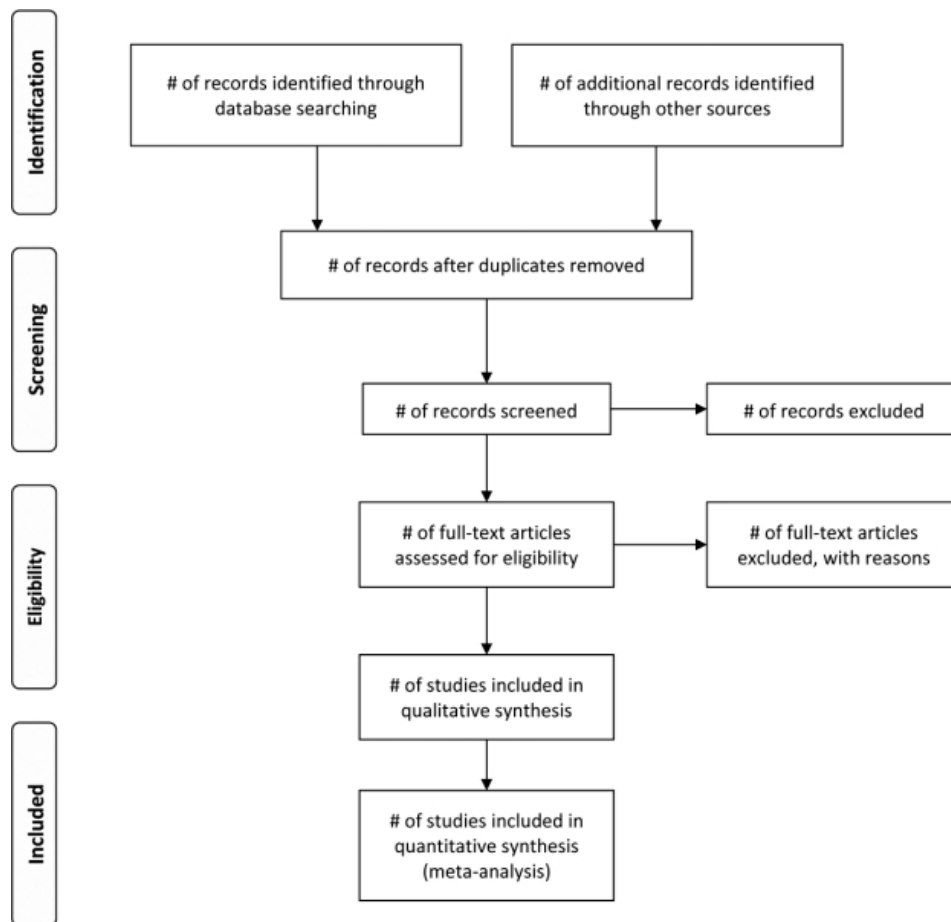
### 3.1. Revisión sistemática y selección de estudios

En primer lugar se realizó una revisión exhaustiva de estudios presentes en bases de datos de acceso público que contuviesen datos relacionados con enfermedades cardiovasculares. Los estudios fueron seleccionados de acuerdo a los criterios establecidos en la declaración PRISMA<sup>27,28</sup>, que determina los estándares en la metodología para las revisiones sistemáticas y metaanálisis (figura 3.1).

Para ello, se definieron las palabras clave de búsqueda, con el fin de recoger el máximo de información posible en relación a cada enfermedad. Las bases de datos seleccionadas fueron *Gene Expression Omnibus* (GEO) y *ArrayExpress*<sup>29</sup>, que contienen diferentes tipos de datos ómicos, siendo los transcriptómicos los más numerosos.

En cuanto a los criterios de selección, se realizó una revisión manual de los estudios obtenidos tras aplicar las palabras clave de “*aortic stenosis*” o “*calcific valve disease*” para la estenosis aórtica y “*coronary artery disease*”, “*ischemic cardiomyopathy*” o “*coronary cardiomyopathy*” para la enfermedad de las arterias coronarias. Seguidamente, se aplicaron los siguientes criterios de selección:

- Los estudios realmente centrados en la estenosis aórtica o la enfermedad arterial coronaria.
- Especie humana.
- Tecnologías transcriptómicas de *microarrays* y RNA-Seq.
- Inclusión de la variable sexo y edad de los pacientes.
- Al menos, 10 muestras por condición (caso / control) y preferiblemente 3 por sexo/condición (mujer caso/control y hombre caso/control).
- Tipo de muestra: tejido cardíaco o sangre periférica.



### Year of publication

**Figura 3.1:** Esquema de flujo PRISMA para llevar a cabo la revisión sistemática y metaanálisis.

Se excluyeron por tanto los estudios que tuviesen relación con otras enfermedades cardíacas, como son la miocardiopatía dilatada y la cardiopatía hipertrófica, además de otras enfermedades no cardíacas, como son el Síndrome de Williams, depresión o alcoholismo. También fueron descartados los datos que utilizaban otros datos como los *arrays* no codificantes, estudios epigenéticos y estudios que no tuviesen muestras control.

Para la descarga de datos de los estudios finalmente seleccionados, se utilizó el paquete GEOquery<sup>30</sup>. Debido a la heterogeneidad de los datos y para evitar nuevos sesgos, se procedió a la descarga de los datos ya procesados y normalizados por los investigadores y no los datos brutos (raw data).

## 3.2. Análisis individual de los estudios

Una vez que tenemos descargados los datos del estudio, completamos el procesamiento de los datos cuando fue necesario y se llevó a cabo un análisis exploratorio para disponer de una primera aproximación de la información de los estudios, en cuanto a distribución de las muestras y su relación, además de la heterogeneidad de las mismas.

### 3.2.1. Procesamiento de los datos

Para cada estudio, se realizó un procesamiento de los datos con el fin de estandarizar la nomenclatura y dividir las muestras en cada grupo experimental, además de obtener la información de los genes a partir de los datos de anotación para cada plataforma.

En el caso de la enfermedad de estenosis aórtica, las muestras fueron clasificadas de la siguiente forma:

- **Grupo Control:** Pacientes sanos, divididos por sexo en hombres y mujeres.
- **Grupo Estenosis Aórtica:** Pacientes con la enfermedad cardíaca, divididos también por sexo.

Para la enfermedad de las arterias coronarias y cardiopatía isquémica, se empleó la siguiente nomenclatura:

- **Grupo Control:** Pacientes sanos, divididos por sexo.
- **Grupo Enfermedad Arterial Coronaria:** Pacientes que tienen la enfermedad cardíaca, divididos también por sexo.
- **Grupo Otros:** Pacientes con otras enfermedades cardíacas, como por ejemplo cardiopatía dilatada.

Puesto que cada estudio emplea diferentes plataformas comerciales y modelos, es necesaria la anotación común de las diferentes sondas, para cuantificar la expresión a nivel de gen. Para ello, los códigos identificadores de cada sonda (PROBE\_ID) fueron transformados a un identificador común (ENTREZ\_ID), con el que se puede identificar el gen al que pertenecen. Puesto que en algunos chips de *microarray* varias sondas pueden hibridar con diferentes regiones de un mismo gen, se realizó la mediana de las medidas de expresión de las diferentes sondas para un mismo gen y tener así la información de expresión a nivel de gen, con la función `medianReps`, para cada una de las muestras de los diferentes pacientes, evitando información duplicada.

### 3.2.2. Análisis exploratorio

Una vez que están los datos de expresión anotados, sin información duplicada y con las muestras distribuidas en los diferentes grupos de interés, se realizó un análisis exploratorio de los datos de cada estudio, con el fin de determinar la distribución de la expresión de las muestras, comprobar que no hayan patrones anómalos y evaluar la calidad de los datos del estudio.

Para ello, se realizaron diagramas de caja que permitieron conocer si los grupos de datos debían completar su procesamiento con un reescalado y/o transformación logarítmica.

Además, mediante un análisis de agrupamiento o *clustering* y un análisis de componentes principales (PCA), se evaluó la distribución de las muestras de los grupos experimentales.

### 3.2.3. Análisis de expresión diferencial

Tras la exploración de los datos, se realizó el análisis de expresión diferencial, para determinar qué genes presentaban patrones de expresión diferente entre hombres y mujeres que padecen la enfermedad cardiovascular. Para ello, se realizó una combinación de los factores sexo y enfermedad, planteando en el siguiente contraste:

**(Mujeres caso - Mujeres control) - (Hombres caso - Hombres control)**

Siendo caso la enfermedad cardiovascular de interés.

Para ello, se empleó la función `diffExp` que utiliza el paquete `limma` y que permite evaluar para cada gen un contraste de medias de expresión definido en un modelo lineal, donde la variable respuesta es la expresión génica. El p-valor obtenido se ajustó por el método de Benjamini-Hochberg<sup>31</sup> (BH).

### 3.2.4. Análisis de enriquecimiento funcional

Para identificar las funciones moleculares que están alteradas por la enfermedad cardiovascular en cada sexo, se realizó un análisis de enriquecimiento funcional mediante GSEA a partir de los resultados de la expresión diferencial, obtenidos en el apartado anterior.

La función utilizada para llevar a cabo el enriquecimiento funcional fue `doGSEA`, que utiliza la librería `mdgsa`<sup>32</sup>. Para obtener la anotación funcional se utilizó la librería `biomaRt`, donde a partir de los identificadores Entrez de los genes, se obtuvo la anotación para las diferentes ontologías (términos GO: proceso biológico BP, función molecular MF y componente celular CC). Se realizó una propagación de la anotación de los términos GO, de modo que los genes anotados con un término GO heredan la información de sus ancestros, puesto que la información mantiene una estructura jerárquica.

Además de los términos GO, también se realizó un análisis de enriquecimiento funcional con la información disponible en bases de datos sobre rutas de señalización como son Reactome (de acceso libre) y KEGG (*Kyoto Encyclopedia of Genes and Genomes*). En el caso de esta última base de datos, se utilizó la anotación disponible con el paquete `org.Hs.eg.db`.

Se realizó un filtrado de las funciones con un grupo de genes muy reducido (funciones muy específicas, con un número de genes inferior a 10) y los grupos de genes muy grandes (funciones muy generales, con un número de genes superior a 500).

De esta forma, con los resultados obtenidos en el apartado de expresión diferencial y la anotación funcional se llevó a cabo el análisis de enriquecimiento funcional. Se consideraron significativas las funciones enriquecidas con un p-valor ajustado inferior a 0.05.

### 3.2.5. Análisis de rutas de señalización

Para finalizar el análisis individual de cada estudio, se llevó a cabo un análisis de rutas de señalización. Hipathia<sup>22,33</sup> es una herramienta basada en métodos CCAA (*Canonical Circuit Activity Analysis*) que permite el análisis de actividad de rutas de señalización, a partir de los datos transcriptómicos y los conocimientos sobre la transducción de la señal. El método mencionado anteriormente está basado en un algoritmo iterativo que es capaz de calcular la intensidad de señal que circula a través de los nodos de una red, a partir de los niveles de expresión de cada gen (nodos) y la intensidad de la señal que llega a este nodo. Al final se estima la señal final que llega a los nodos efectores, que tendrán asociados una función determinada.

Las rutas que están incluidas en Hipathia provienen de la base de datos KEGG. Puesto que una ruta puede estar asociada a diferentes nodos efectores, Hipathia divide cada ruta en diferentes subrutas, en los que están incluidos los nodos efectores y todos aquellos nodos asociados a él. De esta forma, se lleva a cabo un análisis de cada subruta.

Para ello, previamente se traducen los genes que participan a su identificador Entrez. Seguidamente se lleva a cabo una normalización de los valores de expresión entre 0 y 1 y finalmente se obtienen los valores de activación para cada subruta a partir de los genes presentes en el estudio.

Una vez que se tienen los valores de activación de cada ruta, se procede a realizar el contraste de interés, con la misma función empleada para el análisis de expresión diferencial (diffExp) que utiliza el paquete limma, obteniendo de esta forma las rutas que están diferencialmente activadas. Se obtuvo la cuantificación de la activación para los términos GO y Uniprot.

## 3.3. Metaanálisis:

Finalmente, se llevó a cabo un metaanálisis a partir de los datos de expresión diferencial de cada estudio individual, para determinar de esta forma los genes que están diferencialmente expresados en el conjunto de todos los estudios seleccionados de una

enfermedad y por otro lado, también se llevó a cabo un metaanálisis de las funciones alteradas a partir de los resultados obtenidos en el enriquecimiento funcional.

### 3.3.1. Metaanálisis de Genes

Como se ha comentado anteriormente, los datos de partida para realizar el metaanálisis de genes son los resultados individuales de la expresión diferencial. Se evaluaron los logaritmos de *fold change* (logFC) que determinan la medida del efecto y el error estándar (SE), que mide la variabilidad de la medida del efecto, creando así 2 matrices. Las filas representan los genes y las columnas el estudio al que pertenecen.

#### **Determinación de la medida combinada del efecto:**

A partir de estas matrices, se realizó el metaanálisis para cada gen presente en los estudios. Generalmente en el metaanálisis primero se evalúa la heterogeneidad de los estudios y en función de esta información, se decide el modelo de estimación de la variabilidad del efecto, pero debido a la gran cantidad de genes presentes en los estudios, se decidió valorar cual es el modelo más apropiado para determinar la variabilidad *a posteriori*. Esta heterogeneidad hace referencia a la variabilidad de los efectos evaluados debido a la diversidad de los datos y a la metodología en los diferentes estudios.

Para realizar el metaanálisis se empleó el paquete *metafor* de R<sup>34</sup>. Se utilizó el método de efectos fijos (FE) y los métodos de efectos aleatorios de DerSimonian y Laird<sup>35</sup> (DL), Hedges<sup>36</sup> (HE) y Hunter y Schmidt (HS)<sup>37</sup>, para su posterior evaluación.

Finalmente el método de efectos aleatorios seleccionado fue el de DL por ajustarse mejor a la característica de los datos ómicos y a la diversidad de plataformas en los estudios.

Al final del metaanálisis se obtiene una tabla en la que aparecen diferentes estimadores de la medida combinada del efecto de los diferentes genes e indicadores de la heterogeneidad del metaanálisis<sup>15</sup>:

- QE y QEp: Estadístico de contraste y p-valor, respectivamente, del método DL. Determinan la heterogeneidad entre estudios. La hipótesis nula indica homogeneidad entre estudios.
- LogFC: Estimación del efecto combinado de los diferentes estudios. El signo determina en qué grupo es mayor el efecto, siendo positivo si hay sobreexpresión del gen en el primer grupo y negativo en el segundo. El valor determina el tamaño del efecto.
- Intervalo de confianza del LogFC (95%): Generado a partir de la variabilidad estimada por el modelo. Si este intervalo no contiene el 0, el logFc se considera significativo.
- P-valor: Indica el nivel de significación del efecto combinado. Puesto que se realizan múltiples análisis, se realizó la corrección por el método BH.
- r2: Estimación de la heterogeneidad entre estudios.

- I<sup>2</sup>: Indica el porcentaje de variabilidad debida a la heterogeneidad en los estudios. Cuanto menor sea este porcentaje, menos relevante es la heterogeneidad.
- H<sup>2</sup>: Cociente entre la variabilidad total y la variabilidad en el muestreo.

### **Evaluación de la heterogeneidad:**

Como consecuencia de la baja potencia que presenta el test Q del método DL al trabajar con un número reducido de estudios, se ha completado el análisis de heterogeneidad con métodos que resumen la información en forma de gráficos de embudo.

En estos gráficos se evalúa la variabilidad de los diferentes estudios con el fin de determinar sesgos. En un eje X se muestra la magnitud del efecto combinado y en el eje y se presenta la medida de la precisión, como es el error estándar.

### **Representación de resultados:**

Además de representar los resultados del metaanálisis, también se realizaron gráficos de bosque para cada gen, donde se pueden evaluar los resultados globales y el peso del resultado de cada gen en el estudio individual.

En estos gráficos, a la izquierda se indican los estudios que están incluidos en el metaanálisis y a la derecha se muestra la estimación de la medida resumen individual de cada uno, con un intervalo de confianza del 95%. En la parte central de la figura se muestra la medida del efecto en recuadros negros, con su intervalo de confianza. En la parte inferior central se muestra el resultado del metaanálisis. La línea central representa el efecto nulo y su desplazamiento hacia uno u otro lado representa la sobreexpresión en un grupo u otro. La anchura determina el intervalo de confianza. También se muestra el modelo de análisis empleado en la parte inferior izquierda (DL) y a la derecha el valor estimado y su intervalo de confianza (figura 3.2).

441178

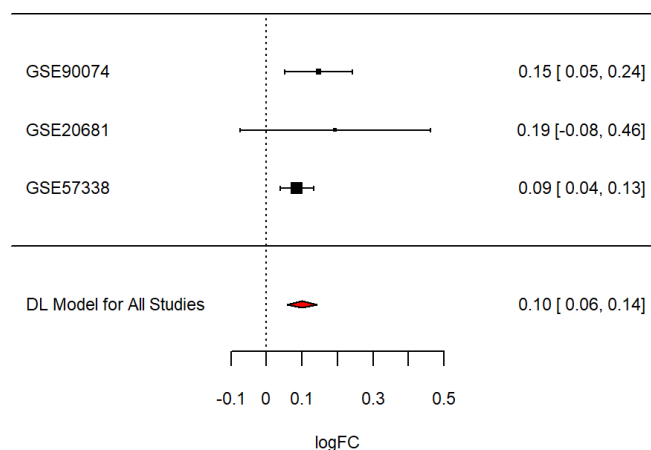


Figura 3.2: Ejemplo de diagrama de bosque.

### 3.3.2. Metaanálisis de funciones

Para realizar el metaanálisis de funciones, se crearon también 2 matrices, pero en lugar de utilizar el  $\log FC$  y el error estándar, se utilizaron los logaritmos de *odds ratios* (LOR) como valor de medida del efecto y la desviación estándar (SD) de los LOR para cada una de las funciones evaluadas, procedentes de los resultados del GSEA de los estudios individuales. De esta forma se evaluaron las funciones con la anotación GO con sus 3 ontologías (procesos biológicos, componentes celulares y funciones moleculares), las rutas con anotación KEGG y finalmente los términos de la HPO.

Se procedió con la misma metodología descrita en el apartado anterior. En este paso los datos de partida son las matrices con los LOR y las SD.

En los resultados del metaanálisis también aparecen los diferentes estimadores indicados anteriormente, junto con el LOR para cada función/ruta como resultado del metaanálisis, además de los gráficos de embudo para determinar la heterogeneidad en los resultados y los gráficos de bosque que resumen los resultados del metaanálisis. Un LOR positivo indica que hay una sobrerrepresentación de esa función/ruta en el grupo de mujeres con la enfermedad cardíaca en relación al grupo de hombres con la enfermedad y viceversa.

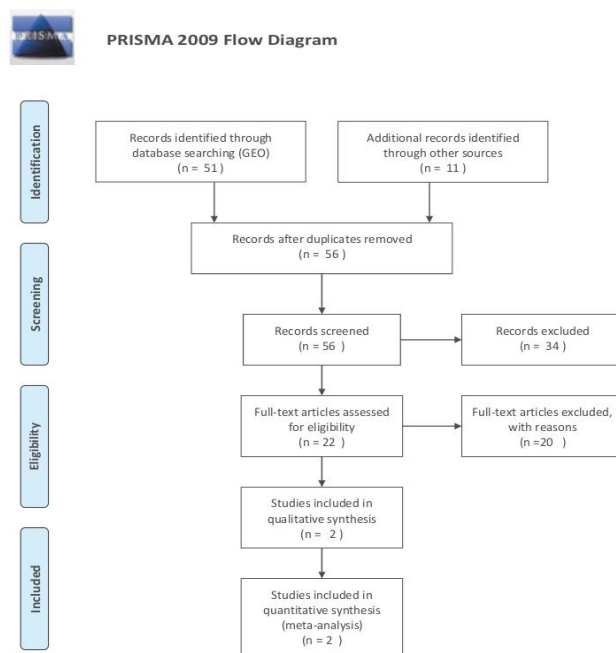
## 4. Resultados

A continuación se muestran los resultados obtenidos tras aplicar la metodología descrita.

### 4.1. Revisión Sistemática y selección de estudios

La búsqueda para los estudios de estenosis aórtica se realizó durante el mes de Julio de 2019 en las bases de datos de GEO y ArrayExpress. Se consideraron las palabras clave “*aortic stenosis*” o “*calcific valve disease*” y organismo humano, identificándose un total de 51 estudios en GEO y 11 en ArrayExpress.

Eliminando los duplicados y aplicando los criterios de inclusión indicados en el apartado anterior, finalmente se seleccionaron 2 estudios. El proceso completo de selección siguiendo el procedimiento PRISMA se muestra en la figura 4.1.



From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(7): e1000097. doi:10.1371/journal.pmed1000097

For more information, visit [www.prisma-statement.org](http://www.prisma-statement.org).

**Figura 4.1:** Proceso de selección para los estudios de Estenosis Aórtica, siguiendo la metodología descrita en la declaración PRISMA.

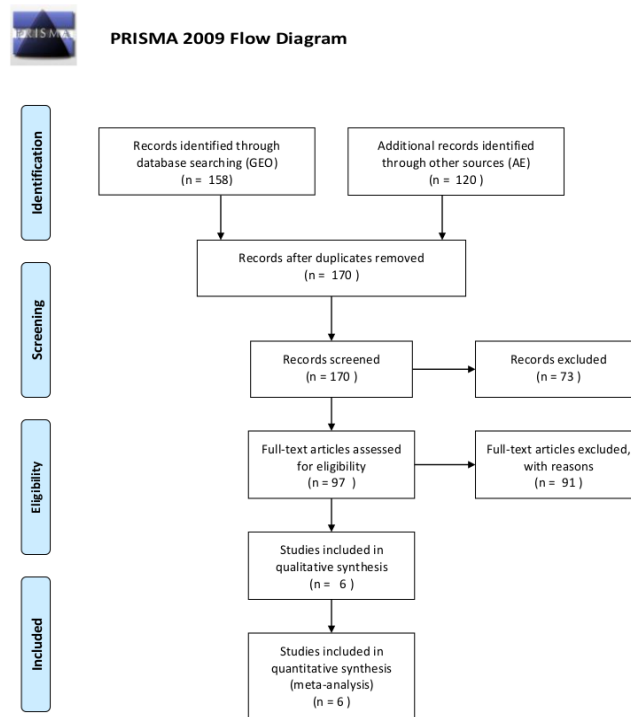
Los estudios seleccionados fueron los siguientes:

| Estudio     | Plataforma                                  | Artículo    |
|-------------|---|-------------|
| GSE10161    | Affymetrix Human Genome U133A Array         | 38          |
| E-MEXP-3684 | Affymetrix GeneChip Human Gene 1.0 ST Array | No presenta |

**Tabla 4.1:** Estudios seleccionados para la enfermedad de estenosis aórtica, se indica la plataforma y publicación asociada (en el caso).

En cuanto a la búsqueda de estudios para enfermedad arterial coronaria, realizada también en Julio de 2019 en los repositorios públicos GEO y ArrayExpress, con los términos “*ischemic cardiomyopathy*” o “*coronary cardiomyopathy*” o “*coronary artery disease*”, se detectaron 158 trabajos en GEO y 120 en ArrayExpress.

Tras eliminar duplicados y aplicar los criterios de inclusión, fueron seleccionados un total de 6 estudios. El proceso completo de selección se detalla en la figura 4.2.



From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(7): e1000097. doi:10.1371/journal.pmed1000097

For more information, visit [www.prisma-statement.org](http://www.prisma-statement.org).

**Figura 4.2:** Proceso de selección para los estudios de enfermedad coronaria arterial, según la declaración PRISMA.

Los estudios finalmente seleccionados fueron los siguientes:

| Estudio   | Plataforma  | Artículo |
|-----------|---|----------|
| GSE120774 | Affymetrix Human Gene 1.0 ST Array                        | 39       |
| GSE90074  | Agilent-014850 Whole Human Genome Microarray 4x44K G4112F | 40       |
| GSE20681  | Agilent-014850 Whole Human Genome Microarray 4x44K G4112F | 41       |
| GSE12288  | Affymetrix Human Genome U133A Array                       | 42       |
| GSE57338  | Affymetrix Human Gene 1.1 ST Array                        | 43       |
| GSE116250 | Illumina HiSeq 2500                                       | 44       |

**Tabla 4.2:** Selección de estudios para la enfermedad de las arterias coronarias. Se indica la plataforma y la publicación asociada (en su caso).

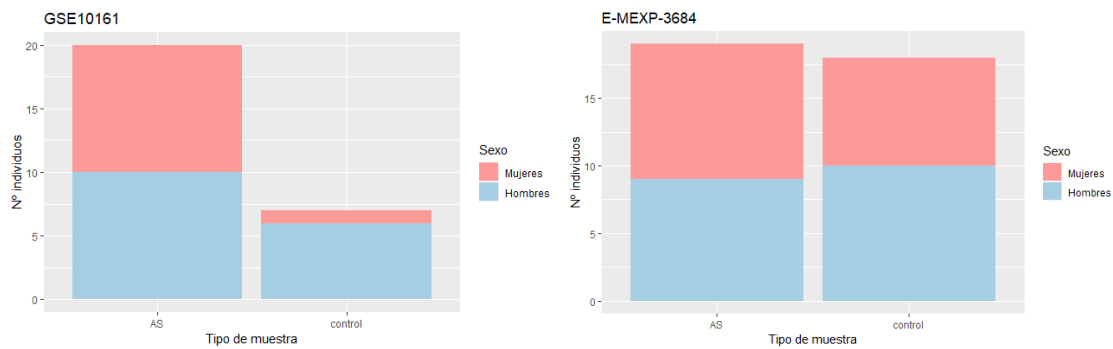
## 4.2. Análisis bioinformático de estudios individuales

### 4.2.1. Procesamiento de los datos

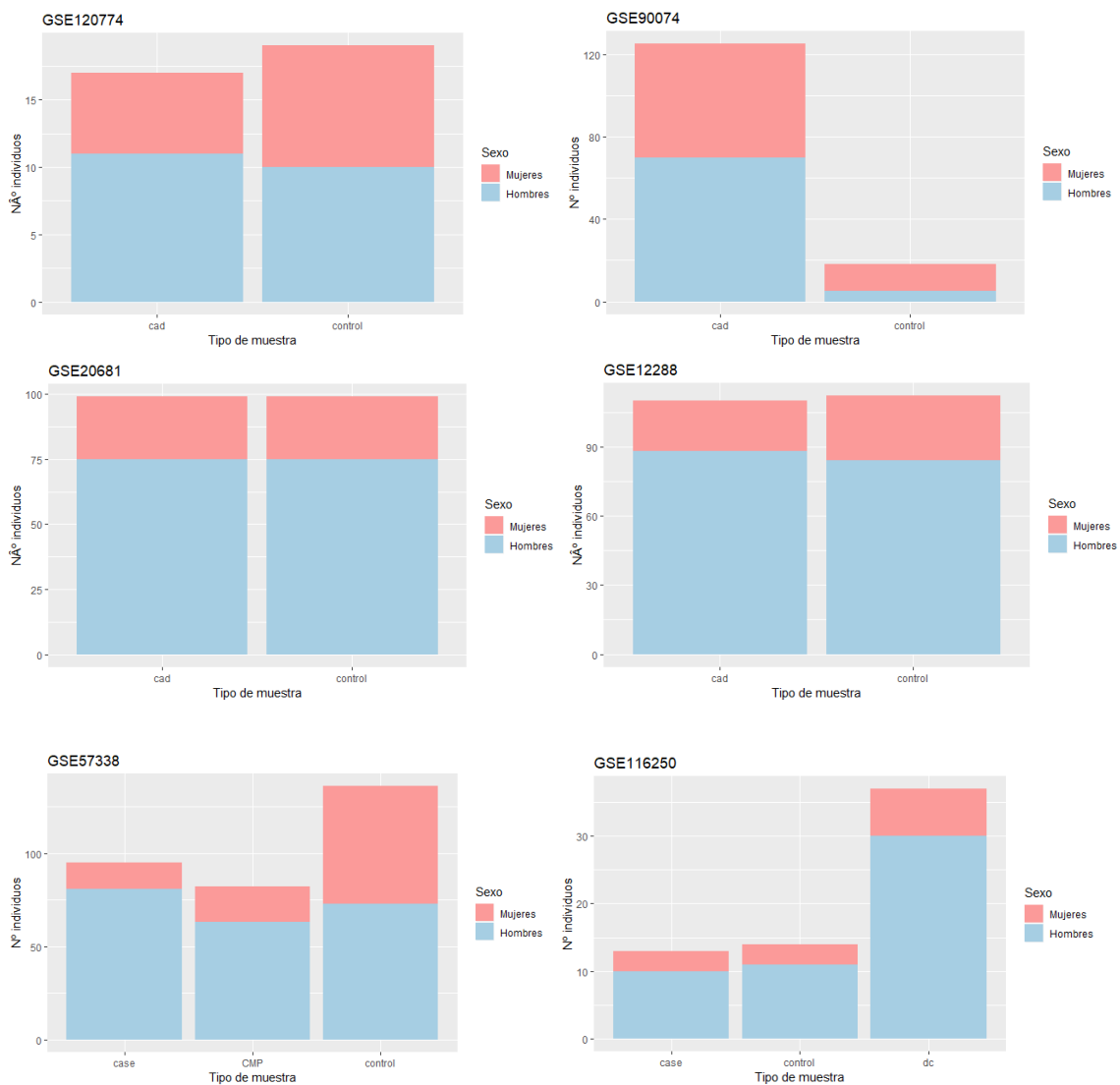
Una vez que los datos de cada estudio fueron descargados, se estandarizó la nomenclatura de los grupos experimentales de acuerdo a lo establecido en los métodos. A continuación, se exploraron los datos completando el procesamiento de los datos normalizados en los estudios GSE10161, GSE12288 y GSE57338, donde se llevó a cabo una transformación logarítmica en base 2, para conseguir una escala homogénea en el conjunto de los datos de los estudios.

En el caso del estudio GSE116250 que procedía de un estudio de RNA-Seq, los datos normalizados aportados por los investigadores estaban expresados en RPKM (Lecturas por kilobase, por millón de lecturas). Se realizó un filtrado de los genes no tenían ninguna lectura y se también se realizó la transformación logarítmica en base 2.

La distribución de muestras por sexo y por grupo fue la siguiente para cada estudio:



**Figura 4.3:** Distribución de las muestras para los estudios de estenosis aórtica. AE: estenosis aórtica y control: pacientes sin la enfermedad.



**Figura 4.4:** Distribución de muestras por sexo y grupo para la enfermedad de las arterias coronarias. Case/Cad: son pacientes con la enfermedad, control: grupo sin la enfermedad y DC/CMP: grupo con otras alteraciones cardíacas (no utilizadas en el análisis).

A continuación se describe cada uno de los estudios:

## Estenosis Aórtica

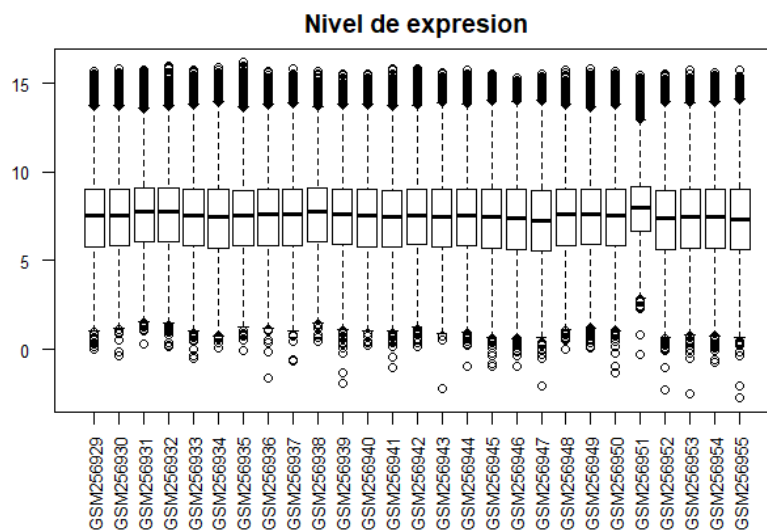
### GSE10161

En este estudio se identifica el papel de la Osteoglicina como regulador de la masa ventricular derecha mediante estudios genómicos y de cómo varía su nivel de expresión dependiendo de condiciones como es el caso de la estenosis aórtica. A continuación se muestra una distribución de las muestras:

|                   | Hombres | Mujeres | Totales |
|-------------------|---------|---------|---------|
| Estenosis Aórtica | 10      | 10      | 20      |
| Controles         | 6       | 1       | 7       |
| Totales           | 16      | 11      | 27      |

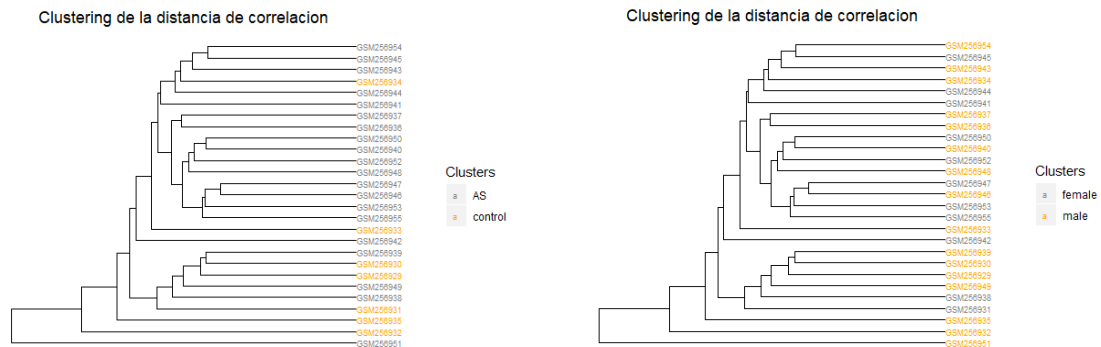
**Tabla 4.3:** Distribución de las muestras por sexo y condición en el estudio GSE120774

Los niveles de expresión fueron los siguientes:

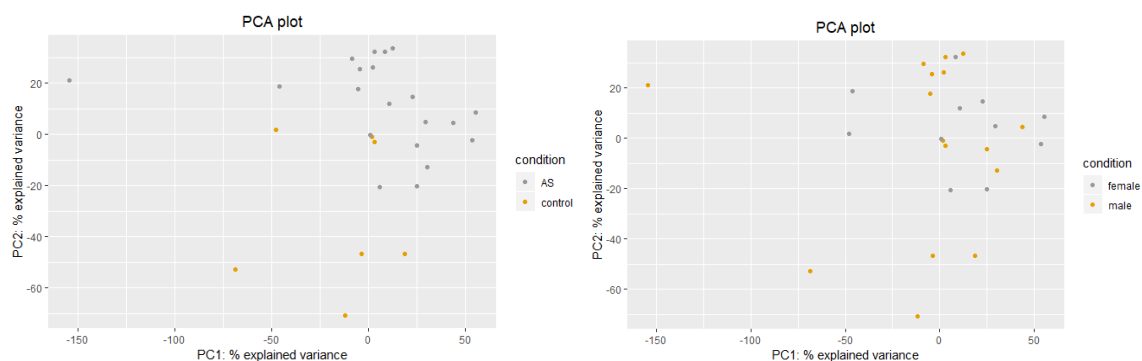


**Figura 4.5:** Niveles de expresión de las diferentes muestras del estudio GSE120774. Con este diagrama de caja se observa como los datos están normalizados.

## Distribución de las muestras



**Figura 4.6:** Distribución de las muestras. A la izquierda, separación por condición, en gris pacientes con estenosis aórtica y en amarillo, pacientes control. A la derecha, separación por sexo, en amarillo hombres y en gris mujeres.



**Figura 4.7:** Análisis de componentes principales para el estudio GSE120774. A la izquierda separación por enfermedad y a la derecha por sexo.

Tanto en el *clustering* como en el PCA se observa como hay una mayor separación de las muestras por la condición que con respecto al sexo.

Puesto que en este estudio solo había 1 muestra de mujer control, finalmente se descartó este estudio a la hora de realizar los análisis posteriores.

### E-MEXP-3684

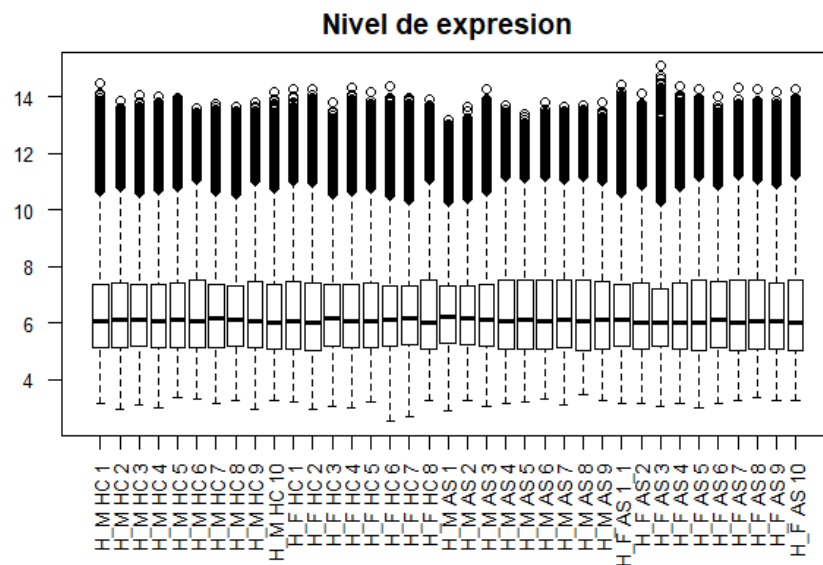
En este estudio se pretende analizar la diferencia de expresión genética entre hombres y mujeres que presentan estenosis aórtica. No tiene publicación asociada.

La distribución de las muestras por condición y sexo fue la siguiente:

|                   | Hombres | Mujeres | Totales |
|-------------------|---------|---------|---------|
| Estenosis Aórtica | 9       | 10      | 19      |
| Controles         | 10      | 8       | 18      |
| Totales           | 19      | 18      | 37      |

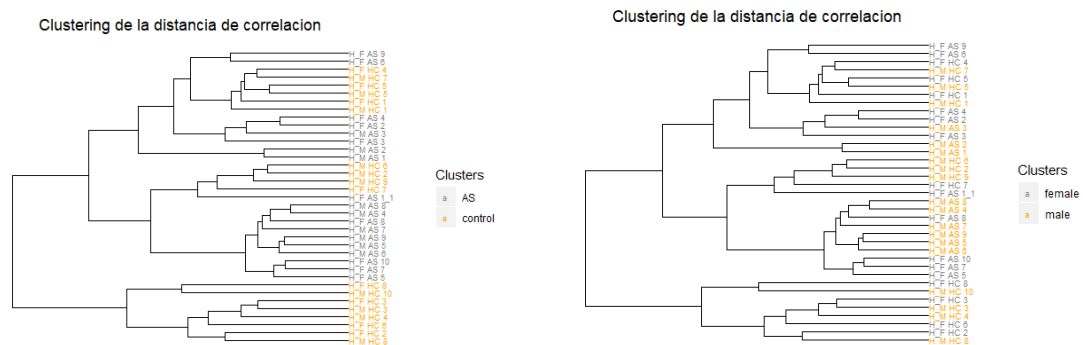
**Tabla 4.4:** Distribución de sexo y condición para el estudio E-MEXP-3684

A continuación se muestra el perfil de expresión de las muestras:

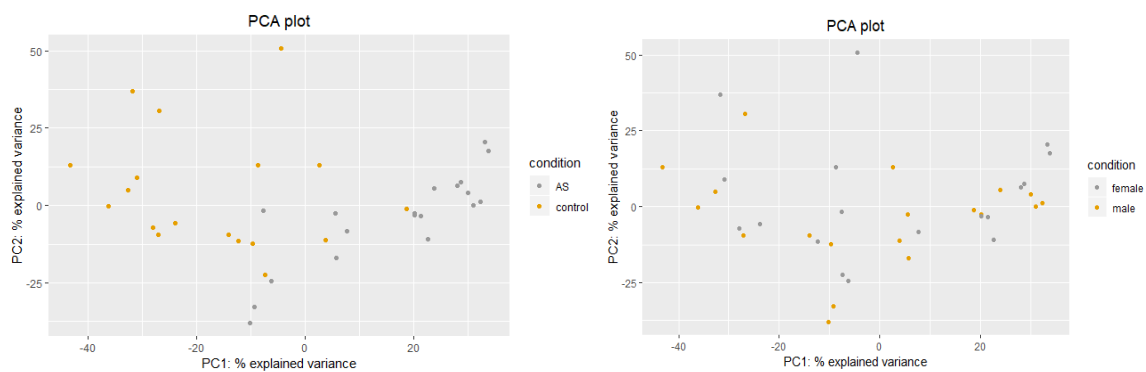


**Figura 4.8:** Nivel de expresión de las diferentes muestras del estudio.

A continuación se muestra la distribución de las muestras:



**Figura 4.9:** Agrupación de las muestras por enfermedad (izquierda) y por sexo (derecha).



**Figura 4.10:** Análisis de componentes principales de las muestras por condición (izquierda) y por sexo (derecha)

## Enfermedad de las Arterias Coronarias

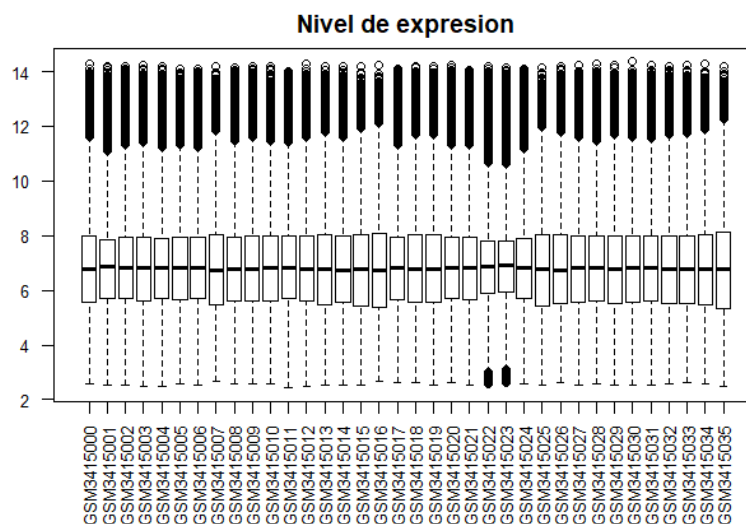
### GSE120774

En el siguiente estudio se analizó el perfil de expresión de tejido adiposo epicárdico en pacientes con la enfermedad de las arterias coronarias con el fin de determinar si hay una mayor expresión de genes relacionados con la inflamación. La distribución de las muestras fue la siguiente:

|           | Hombres | Mujeres | Totales |
|-----------|---------|---------|---------|
| CAD       | 11      | 6       | 17      |
| Controles | 10      | 9       | 19      |
| Totales   | 21      | 15      | 36      |

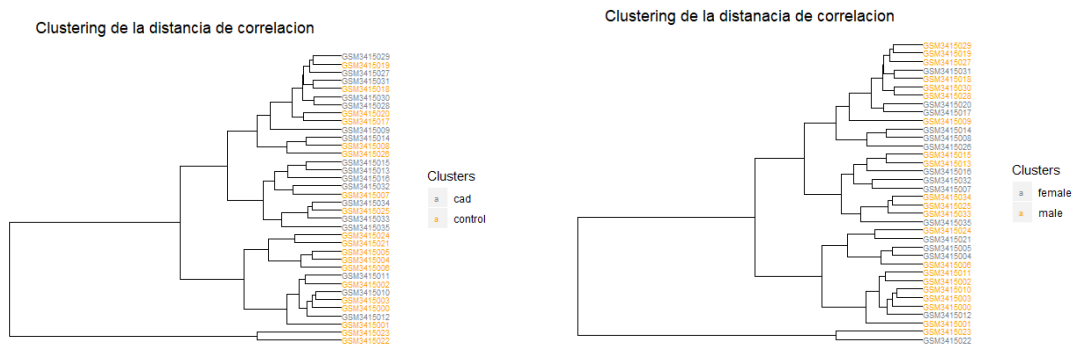
**Tabla 4.5:** Distribución de las muestras para el estudio GSE120774

Seguidamente se muestra el perfil de expresión de las muestras:

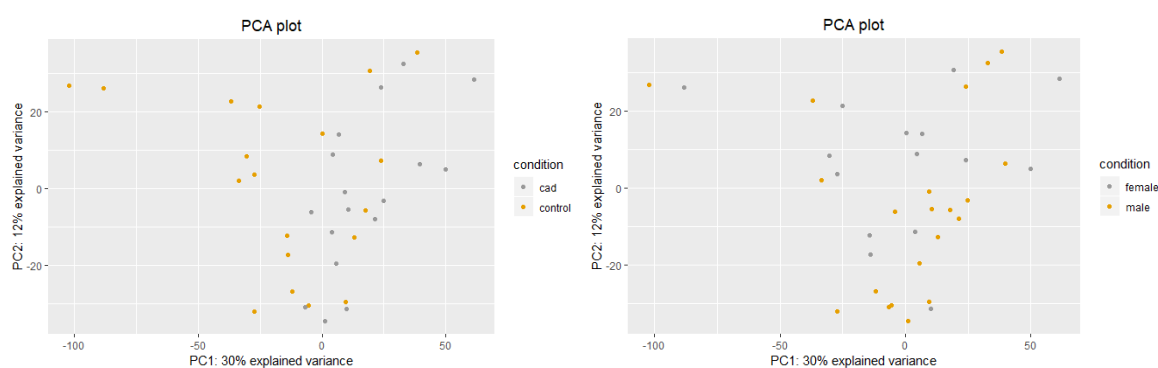


**Figura 4.11:** Perfil de expresión de las muestras del estudio GSE120774.

En cuanto a la distribución de las muestras, fue la siguiente:



**Figura 4.12:** Agrupación de las muestras por enfermedad (izquierda) y sexo (derecha)



**Figura 4.13:** PCA de las las muestras por enfermedad (izquierda) y por sexo(derecha).

Tanto en la agrupación de las muestras (figura) como en los PCA (figura), no se observa una clara separación de las muestras por grupos.

### GSE90074

En este estudio se analiza el perfil de expresión de células mononucleares en pacientes de avanzada edad con aterosclerosis para determinar principalmente el papel de citoquinas proinflamatorias como es CXCL5 en esta enfermedad. Este trabajo es interesante porque además del sexo, aparece información sobre otras alteraciones como hipertensión, diabetes y si el paciente es fumador, aunque esta información no se ha utilizado en el análisis.

La distribución de las muestras fue la siguiente:

|         | Mujeres | Hombres | Total |
|---------|---------|---------|-------|
| CAD     | 55      | 70      | 125   |
| Control | 13      | 5       | 18    |
| Total   | 68      | 75      | 143   |

**Tabla 4.6:** Distribución de las muestras por grupos

Los perfiles de expresión de las muestras fue el siguiente:

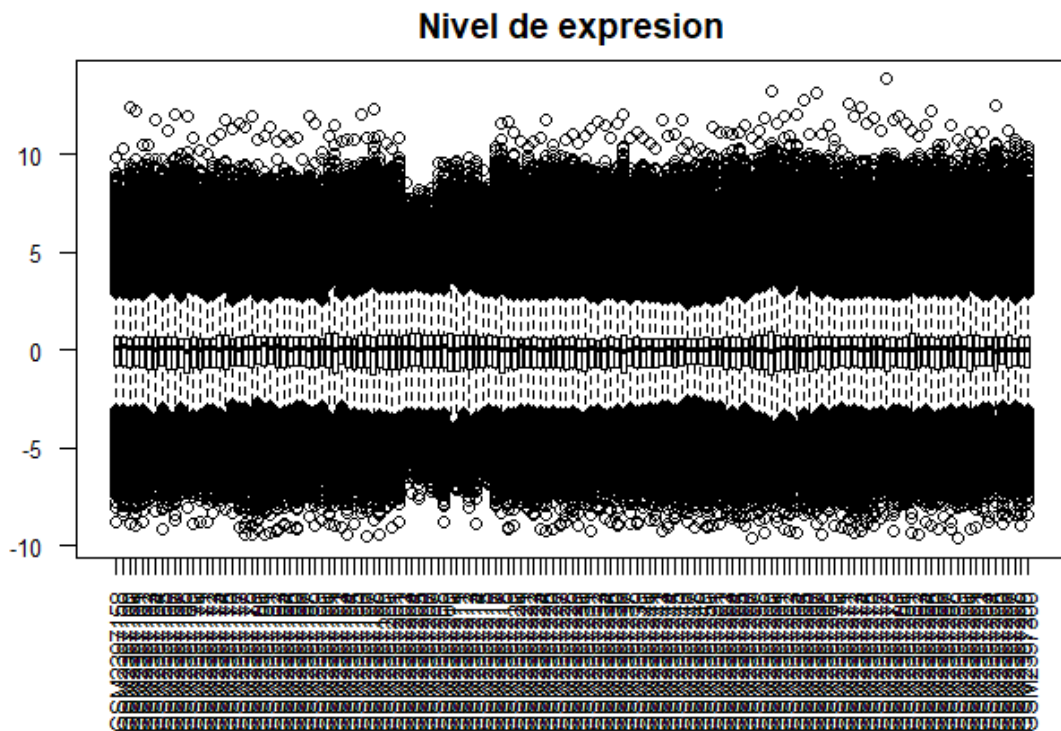


Figura 4.14: Perfil de expresión para el estudio GSE90074.

A continuación se muestra la agrupación de las muestras.

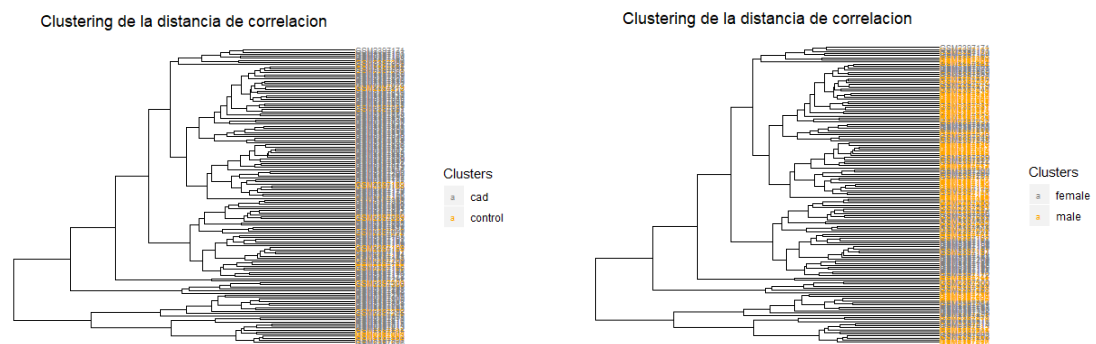
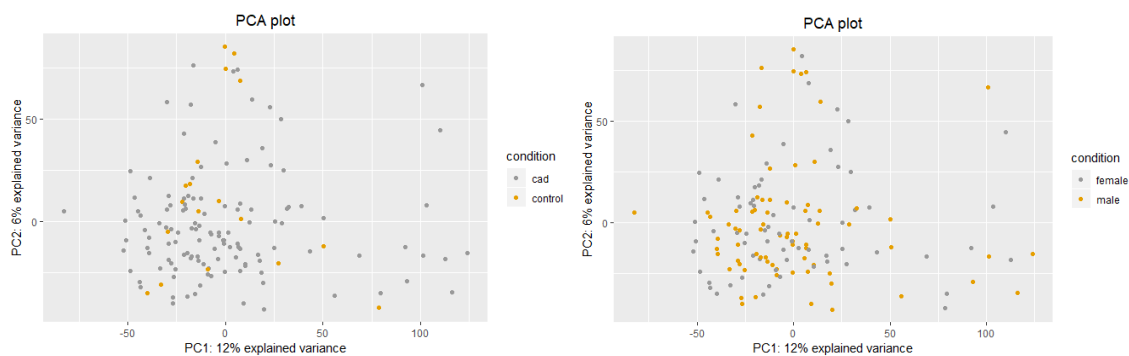


Figura 4.15: Distribución de las muestras por enfermedad (izquierda) y por sexo (derecha).



**Figura 4.16:** PCA del estudio GSE90074 en el que se muestra la agrupación de las muestras por enfermedad (izquierda) y por sexo (derecha).

Como se puede apreciar en las figuras 4.14 y 4.15, no se observa una agrupación clara de las muestras por sexo o enfermedad y en análisis posteriores puede ser que no haya diferencias significativas.

### GSE20681

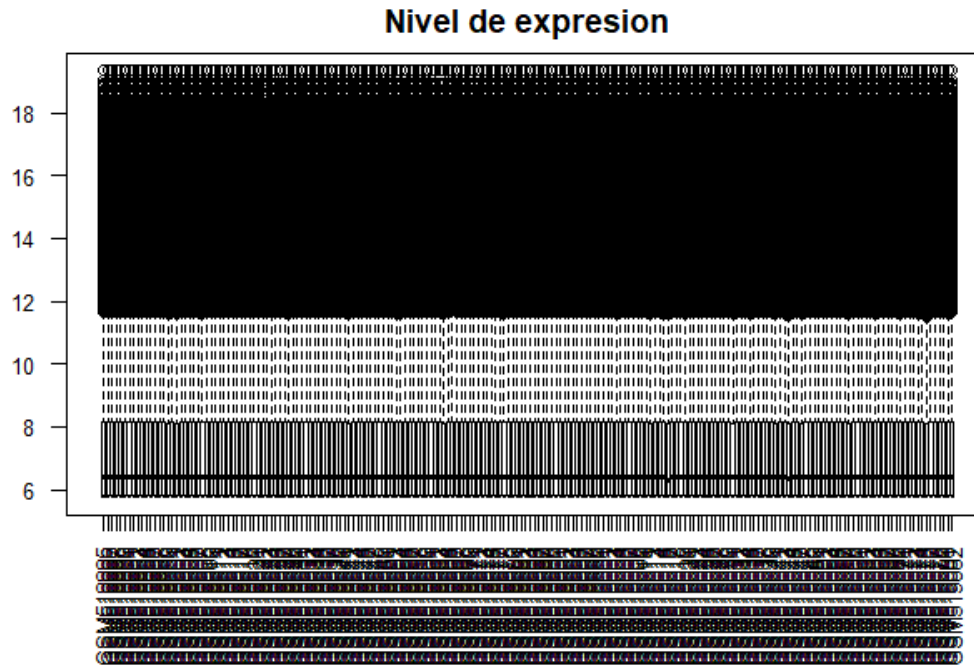
En este estudio se analizó el perfil de expresión de células sanguíneas de pacientes con el fin de encontrar biomarcadores asociados a la enfermedad de las arterias coronarias y desarrollar test no invasivos para detectar la enfermedad.

La distribución de las muestras por sexo y enfermedad fue la siguiente:

|         | Mujeres | Hombres | Total |
|---------|---------|---------|-------|
| CAD     | 24      | 75      | 99    |
| Control | 24      | 75      | 99    |
| Total   | 48      | 150     | 198   |

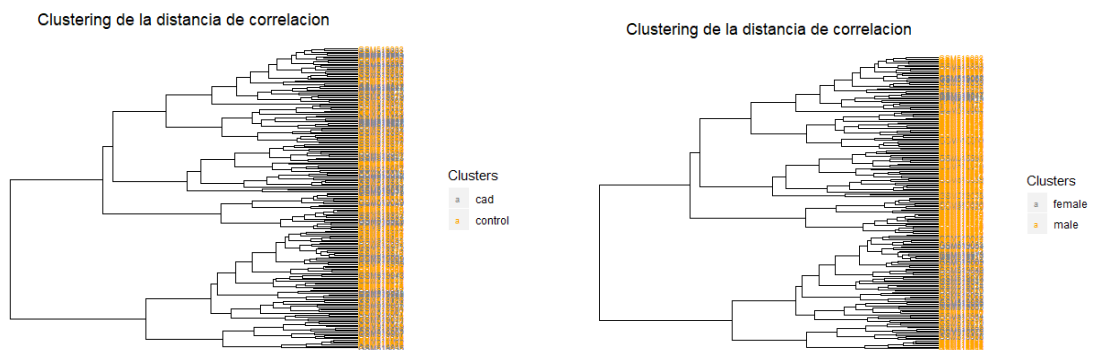
**Tabla 4.7:** Distribución de las muestras por sexo y edad para el estudio GSE2068

Seguidamente se muestra el perfil de expresión de las muestras.

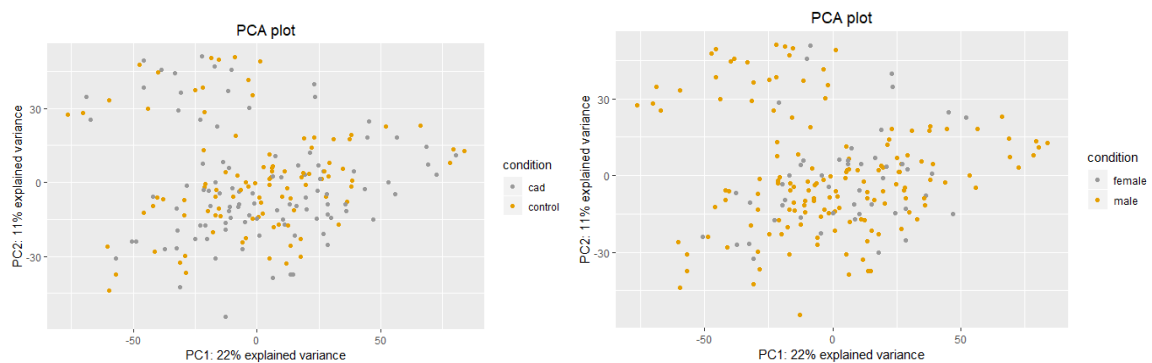


**Figura 4.17:** Perfil de expresión de las muestras del estudio.

A continuación se muestra la agrupación de las muestras del estudio.



**Figura 4.18:** Agrupación de las muestras del estudio por enfermedad (izquierda) y por sexo (derecha).



**Figura 4.19:** PCA del estudio GSE20681 con la distribución por enfermedad (izquierda) y por sexo (derecha).

Como ocurre con el estudio anterior, que también presentaba muchas muestras, no se observa una agrupación clara entre grupos.

### GSE12288

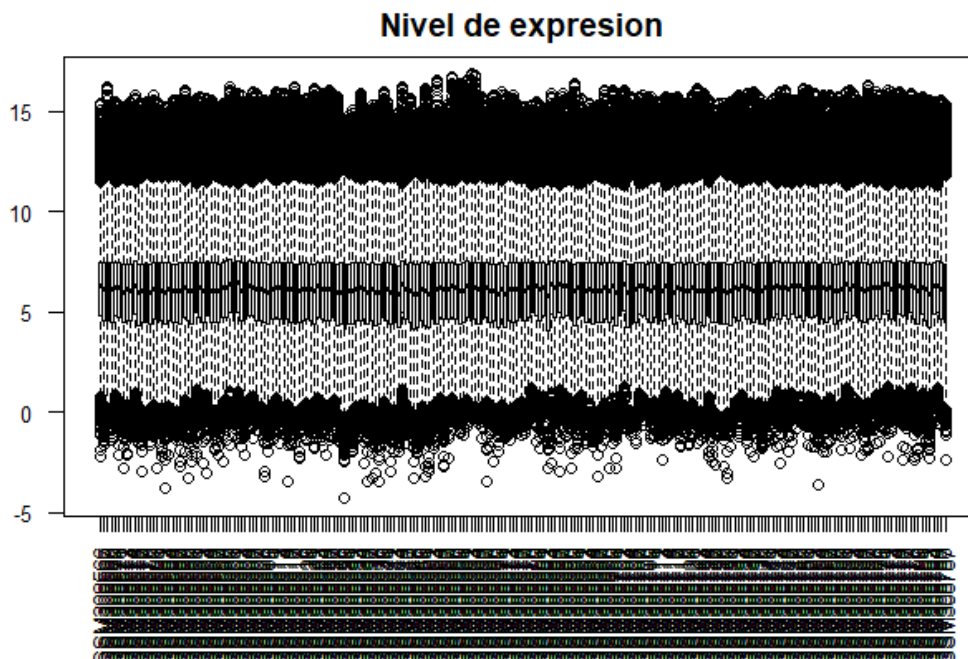
En este estudio se analiza también el perfil de expresión de células sanguíneas (leucocitos) para encontrar genes que presentan expresión diferencial en relación a pacientes con la enfermedad arterial coronaria y pacientes control, para así determinar marcadores de la enfermedad.

La distribución de las muestras fue la siguiente:

|         | Mujeres | Hombres | Total |
|---------|---------|---------|-------|
| CAD     | 22      | 88      | 110   |
| Control | 28      | 84      | 112   |
| Total   | 50      | 172     | 222   |

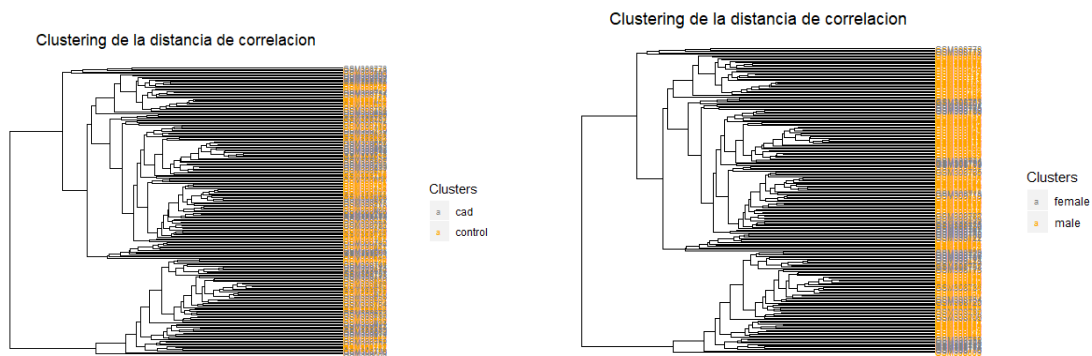
**Tabla 4.8:** Distribución de las muestras en el estudio GSE12288

Los niveles de expresión de las muestras fueron normalizados con una transformación logarítmica en base 2 y fue el siguiente:

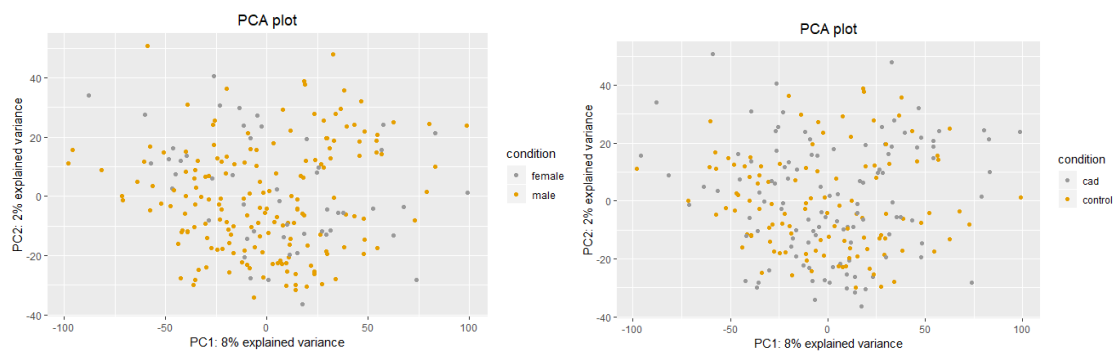


**Figura 4.20:** Perfil de expresión de las muestras del estudio GSE12288.

Seguidamente se muestra la agrupación de las muestras:



**Figura 4.21:** Agrupación por enfermedad (izquierda) y por sexo (derecha).



**Figura 4.22:** PCA del estudio agrupados por enfermedad (izquierda) y por sexo (derecha).

### GSE57338

En este estudio también se utiliza el perfil de expresión para la búsqueda de marcadores de enfermedad cardíaca. Los pacientes, además de enfermedad de las arterias coronarias, también presentaban cardiomiopatía dilatada, aunque estas muestras no fueron analizadas. Fue necesaria también la transformación logarítmica en base 2 de las muestras para equiparlos al resto de estudios.

A continuación se muestra la distribución de las muestras del estudio.

|         | Mujeres | Hombres | Total |
|---------|---------|---------|-------|
| CAD     | 14      | 81      | 95    |
| Control | 63      | 73      | 136   |
| Total   | 77      | 154     | 231   |

**Tabla 4.9:** Distribución de las muestras para el estudio GSE57338. No se incluyeron las muestras de pacientes con otras alteraciones cardíacas.

Seguidamente se muestra el perfil de expresión de las muestras.

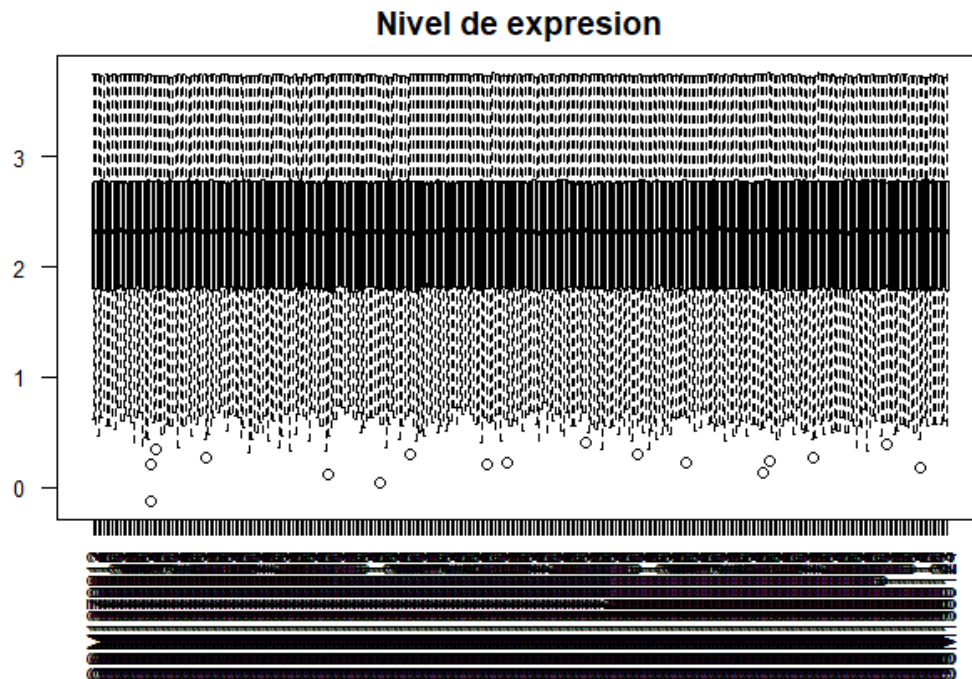


Figura 4.23: Perfil de expresión de las muestras del estudio.

La agrupación de las muestras fue la siguiente:

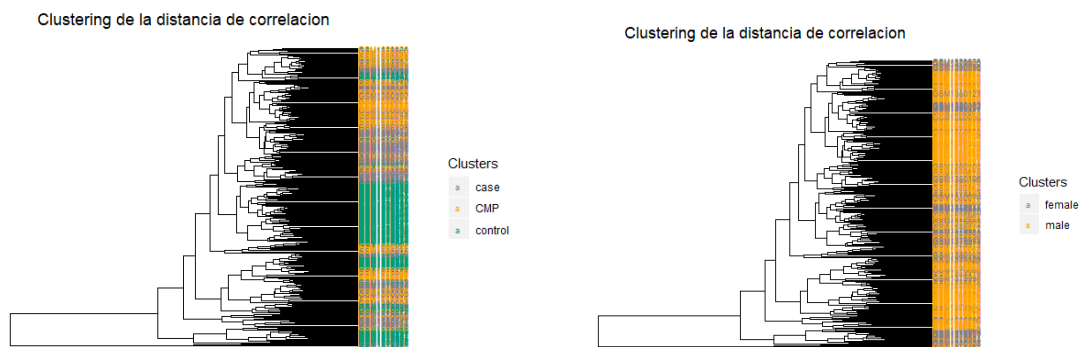
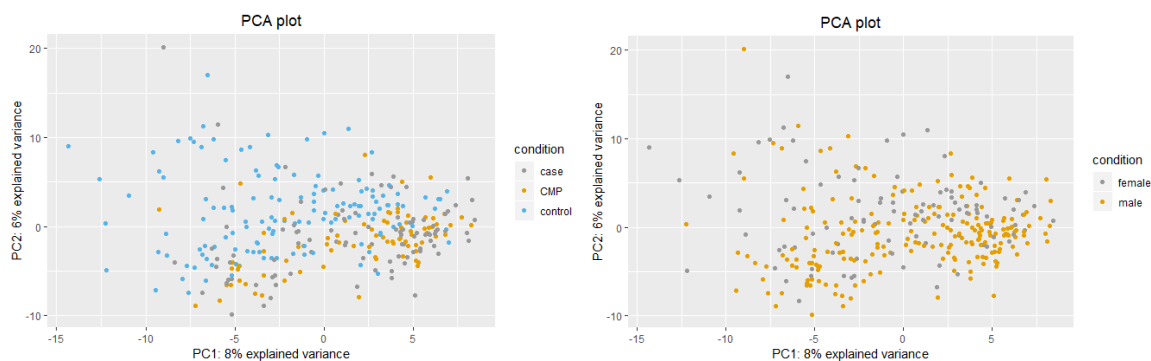


Figura 4.24: Agrupación de las muestras por condición (izquierda) y por sexo (derecha). Por condición, también se muestran los pacientes con otras alteraciones cardíacas (verde) que no son enfermedad arterial coronaria, aunque no se utilizaron para realizar el análisis.



**Figura 4.25:** PCA para las muestras agrupadas por condición (izquierda) y por sexo (derecha). En la agrupación por condición se muestran pacientes con otras alteraciones cardíacas que no son enfermedad arterial coronaria (amarillo).

### GSE116250

En este estudio se utiliza la secuenciación de RNA para determinar los genes que participan en el fallo cardíaco debido a diversas enfermedades como son la cardiopatía isquémica asociada a la enfermedad de las arterias coronarias y a la cardiopatía dilatada. Los datos normalizados aportados por los investigadores de las lecturas estaban expresados en RPKM (lecturas por kilobase, por millón de lecturas). Al ser un dato continuo, se trató como en la metodología anterior para la expresión en *microarrays*. Fueron eliminadas las muestras de pacientes con otras alteraciones cardíacas y se realizó una transformación logarítmica en base 2. También fueron eliminados los genes sin ninguna expresión.

La distribución de las muestras fue la siguiente:

|         | Mujeres | Hombres | Total |
|---------|---------|---------|-------|
| CAD     | 3       | 10      | 13    |
| Control | 3       | 11      | 14    |
| Total   | 6       | 21      | 27    |

**Tabla 4.10:** Distribución de las muestras del estudio GSE116250.

A continuación, se muestra el perfil de expresión de las muestras:

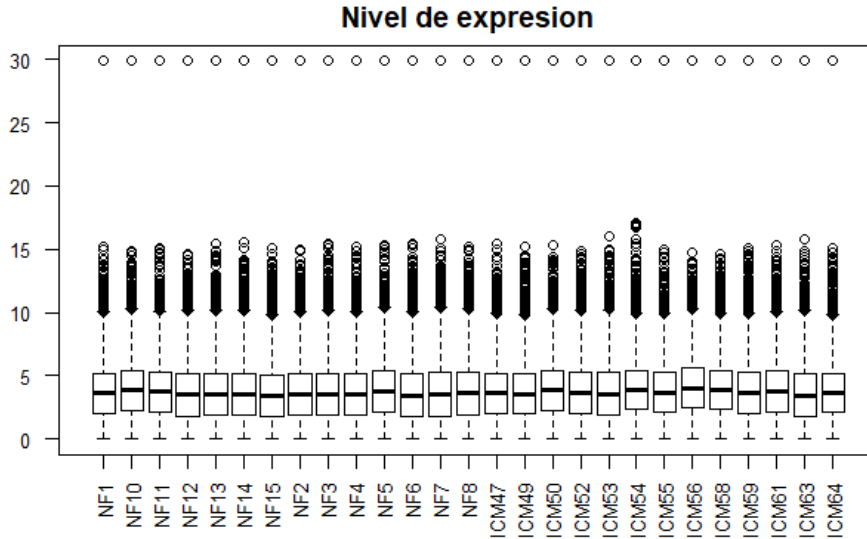


Figura 4.26: Perfil de expresión para el estudio GSE116250.

Finalmente se muestra la agrupación de las muestras del estudio:

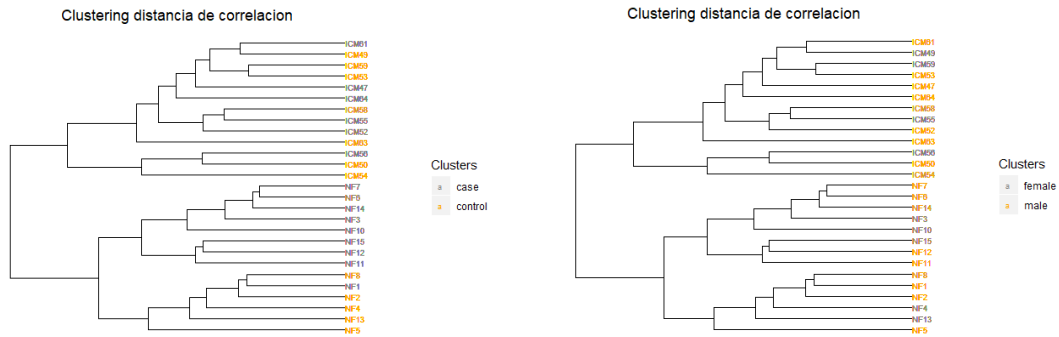


Figura 4.27: Distribución de las muestras por enfermedad (izquierda) y por sexo (derecha).

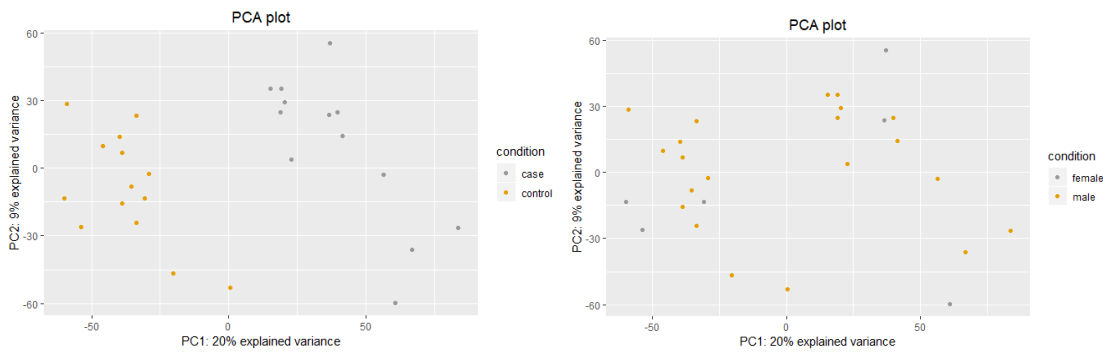


Figura 4.28: PCA agrupadas por enfermedad (izquierda) y por sexo (derecha).

En este caso, la separación de las muestras por enfermedad (figura) es más evidente que por sexo.

#### 4.2.2. Expresión diferencial

En la tablas que se muestran a continuación (tabla 4.11 y tabla 4.12) se resumen los resultados obtenidos tras realizar el análisis de expresión diferencial para el contraste **(Mujeres caso - Mujeres control) - (Hombres caso - Hombres control)**, dónde caso son pacientes con estenosis aórtica o enfermedad de las arterias coronarias según la respectiva tabla. También se llevó a cabo, con fines exploratorios, el contraste entre mujeres caso y mujeres control por un lado y hombres caso y hombres control por el otro lado.

Todos los resultados significativos mostrados presentaron un p-valor ajustado inferior a 0.05, tras aplicar el método BH. Si el logFC fue positivo se indica como *Up*, donde esos genes presentan mayor expresión en mujeres con la enfermedad en relación con los hombres, y *Down* en el caso contrario.

|             |      | (ME - MC) - (HE - HC) | (ME - MC) | (HE - HC) |
|-------------|------|-----------------------|-----------|-----------|
| E-MEXP-3684 | UP   | 0                     | 2016      | 1690      |
|             | DOWN | 0                     | 1784      | 754       |

**Tabla 4.11:** Tabla resumen de los resultados de expresión diferencial para el estudio de estenosis aórtica. ME: mujer con estenosis, MC: mujer control, HE: hombre con estenosis, HC: hombre control.

Como observamos en la tabla 4.11, no se encontró ningún gen significativo (p-valor ajustado inferior a 0.05) para la enfermedad de la estenosis aórtica en el contraste de interés.

|           |      | (ME - MC) - (HE- HC) | (ME - MC) | (HE- HC) |
|-----------|------|----------------------|-----------|----------|
| GSE120774 | UP   | 0                    | 0         | 0        |
|           | DOWN | 0                    | 0         | 0        |
| GSE90074  | UP   | 7                    | 0         | 4        |
|           | DOWN | 1                    | 0         | 11       |
| GSE20681  | UP   | 0                    | 0         | 0        |
|           | DOWN | 0                    | 0         | 0        |
| GSE12288  | UP   | 0                    | 0         | 0        |
|           | DOWN | 0                    | 0         | 0        |
| GSE57338  | UP   | 0                    | 1982      | 1963     |
|           | DOWN | 0                    | 3964      | 4364     |
| GSE116250 | UP   | 11                   | 1124      | 3416     |
|           | DOWN | 10                   | 1034      | 1815     |

**Tabla 4.12:** Resumen de los resultados de expresión diferencial para los estudios de enfermedad de las arterias coronarias. ME: mujer con enfermedad, MC: mujer control, HE: hombre con enfermedad, HC: hombre control.

A continuación se muestran los genes significativos para uno de los estudios de enfermedad de las arterias coronarias (GSE90074).

| ENTREZID | GENENAMES   | logFC  | adj.P.Val |
|----------|---|--------|-----------|
| 203100   | HtrA serine peptidase 4                                 | 1.4663 | 0.00017   |
| 5468     | peroxisome proliferator activated receptor gamma        | 2.1788 | 0.00063   |
| 3778     | potassium calcium-activated channel subfamily M alpha 1 | -0.528 | 0.00405   |

|       |   |        |          |
|-------|---|--------|----------|
| 5654  | HtrA serine peptidase 1                               | 1.2204 | 0.00405  |
| 4257  | microsomal glutathione S-transferase 1                | 0.7636 | 0.01270  |
| 4619  | myosin heavy chain 1                                  | 1.1373 | 0.01663  |
| 30820 | potassium voltage-gated channel interacting protein 1 | 0.4402 | 0.01966  |
| 51700 | cytochrome b5 reductase 2                             | 1.0398 | 0.045663 |

**Tabla 4.14:** Genes significativos del estudio GSE90074

### 4.2.3. Enriquecimiento funcional

Los resultados del análisis GSEA se muestran en las tablas resumen 4.15 y 4.16. Se muestran los resultados para los términos GO: procesos biológicos (BP), componentes celulares (CC) y funciones moleculares (MF), las rutas KEGG y Reactome con un p-valor ajustado por BH inferior a 0.05. También se han realizado representaciones gráficas *Upset* que permiten mostrar el número de funciones y rutas compartidas por los diferentes estudios en el caso de la enfermedad de las arterias coronarias.

#### Estenosis Aórtica

|             | bp  | cc  | mf  | kegg | react | hpo |
|-------------|-----|-----|-----|------|-------|-----|
| E-MEXP-3684 | 424 | 141 | 150 | 68   | 493   | 253 |

**Tabla 4.15:** Tabla resumen del análisis GSEA con las funciones significativas (p-valor ajustado por BH inferior a 0.05) para los estudios de estenosis arterial. Bp: procesos biológicos, cc: componentes celulares, mf: funciones moleculares, kegg: rutas kegg, react: rutas reactome y hpo: funciones fenotípicas.

A continuación se muestra una selección de resultados para cada ontología de la GO. En anexos (tablas a3 - a7) aparecen las tablas con las funciones/rutas más sobrerrepresentadas en mujeres (LOR más positivo) y más sobrerrepresentadas en hombres (LOR más negativo).

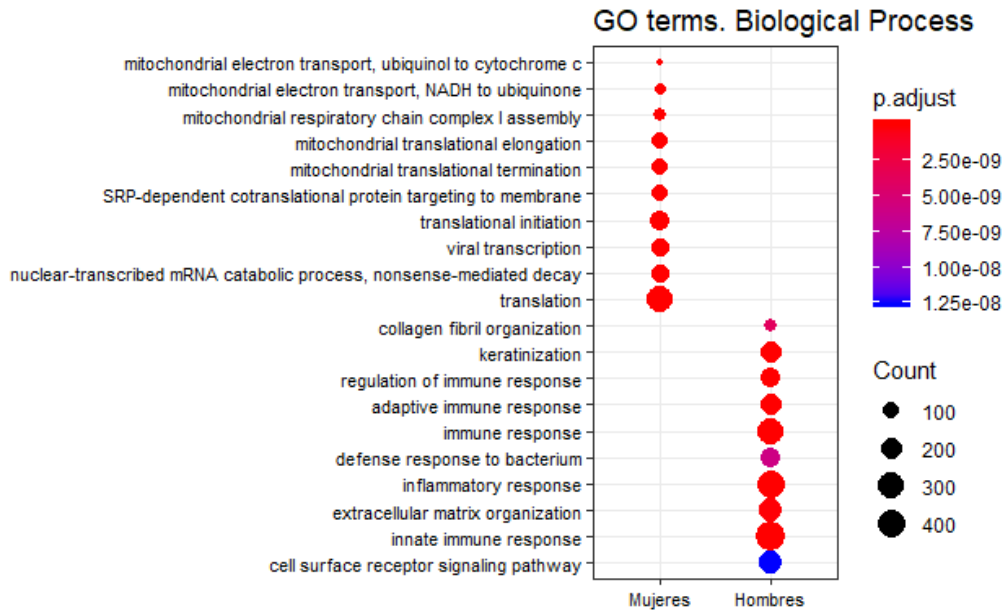


Figura 4.29: Procesos biológicos representativos en mujeres y en hombres para la estenosis aórtica.

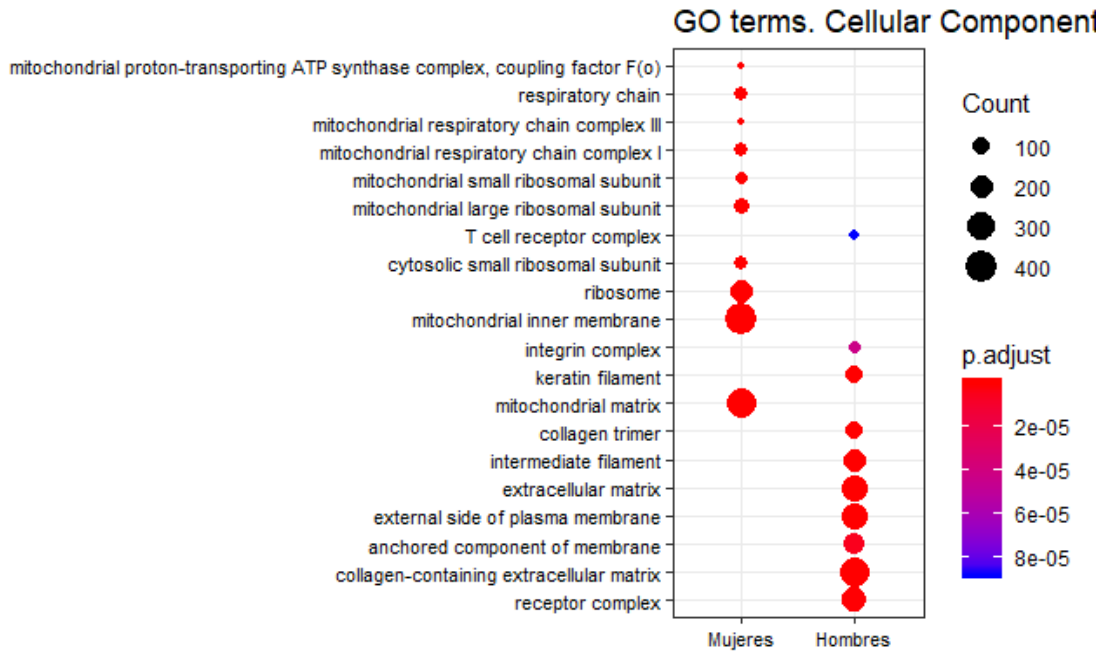
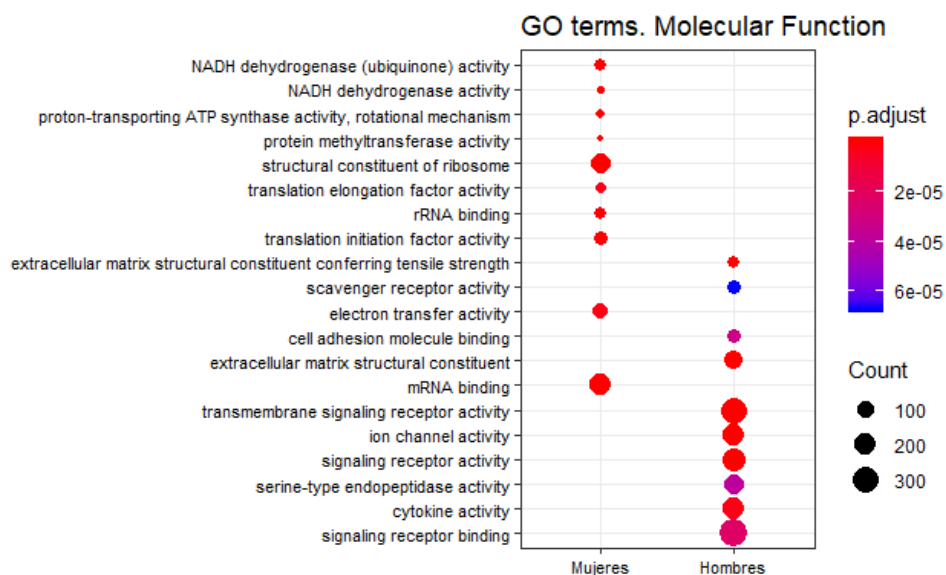


Figura 4.30: Componentes celulares representativos para la estenosis aórtica.



**Figura 4.31:** Funciones moleculares representativas en mujeres y en hombres con estenosis aórtica.

### Enfermedad de las arterias coronarias

|           | bp   | cc  | mf  | kegg | react | hpo |
|-----------|------|-----|-----|------|-------|-----|
| GSE120774 | 1192 | 157 | 210 | 210  | 275   | 608 |
| GSE90074  | 149  | 46  | 10  | 27   | 32    | 1   |
| GSE20681  | 423  | 191 | 117 | 25   | 285   | 10  |
| GSE1288   | 295  | 80  | 34  | 37   | 275   | 12  |
| GSE57338  | 937  | 159 | 84  | 49   | 153   | 659 |
| GSE116250 | 281  | 95  | 74  | 28   | 37    | 98  |

**Tabla 4.16:** Tabla resumen del análisis GSEA con las funciones con un p-valor ajustado por BH inferior a 0.05 para los estudios de la enfermedad de las arterias coronarias.

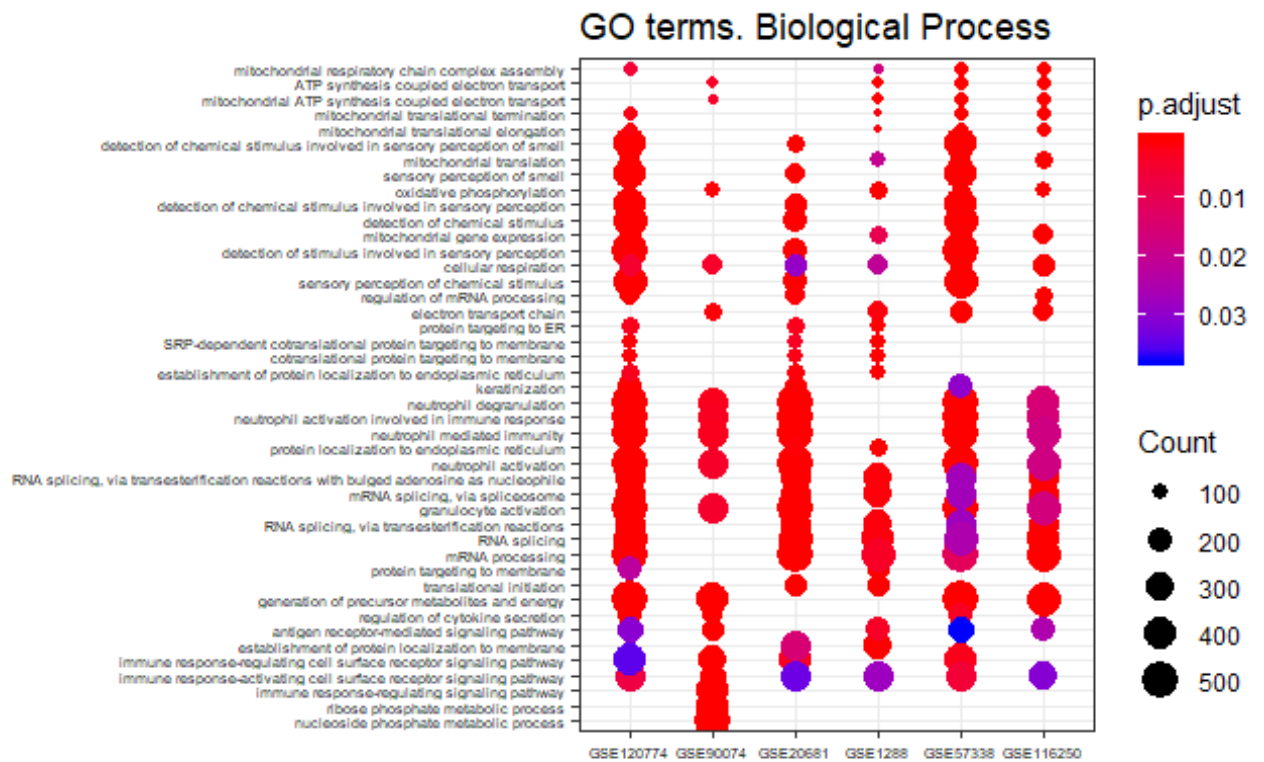


Figura 4.31: Procesos biológicos descritos para la enfermedad de las arterias coronarias.

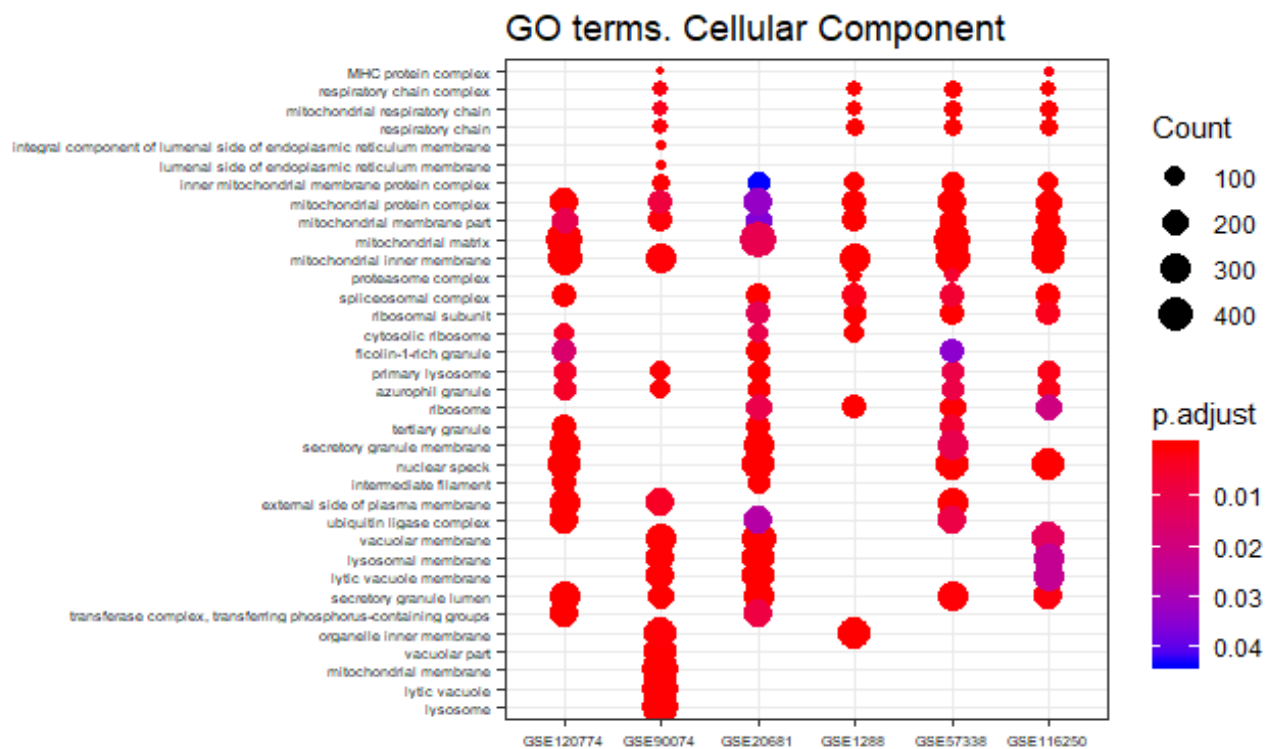
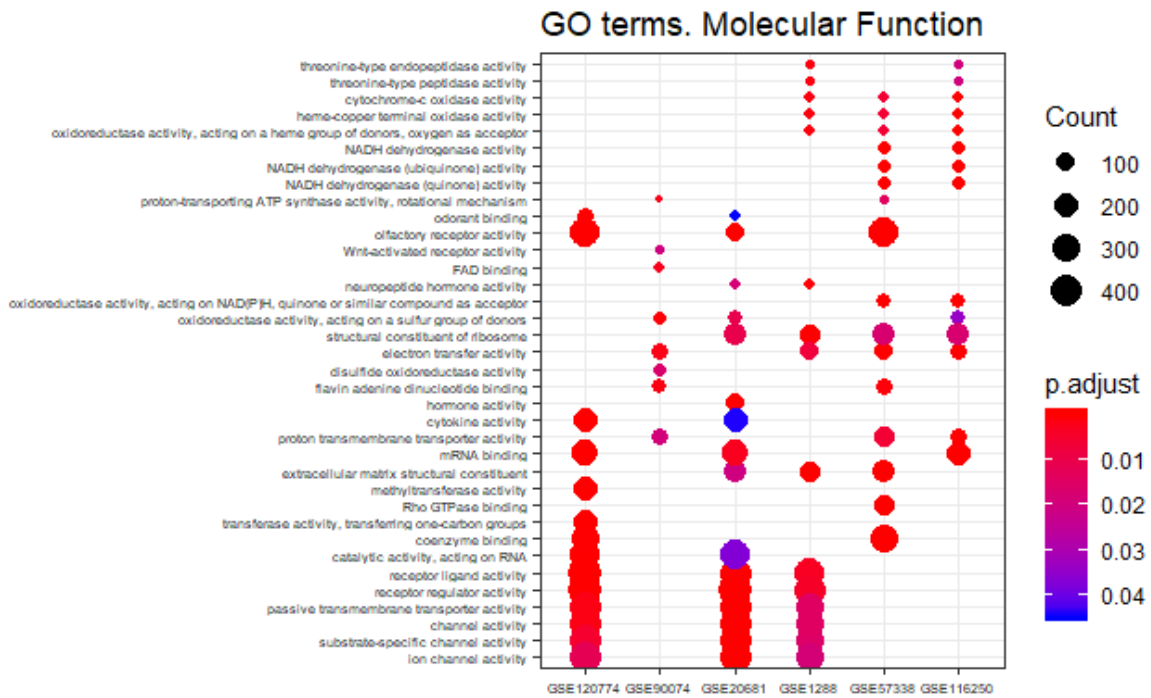
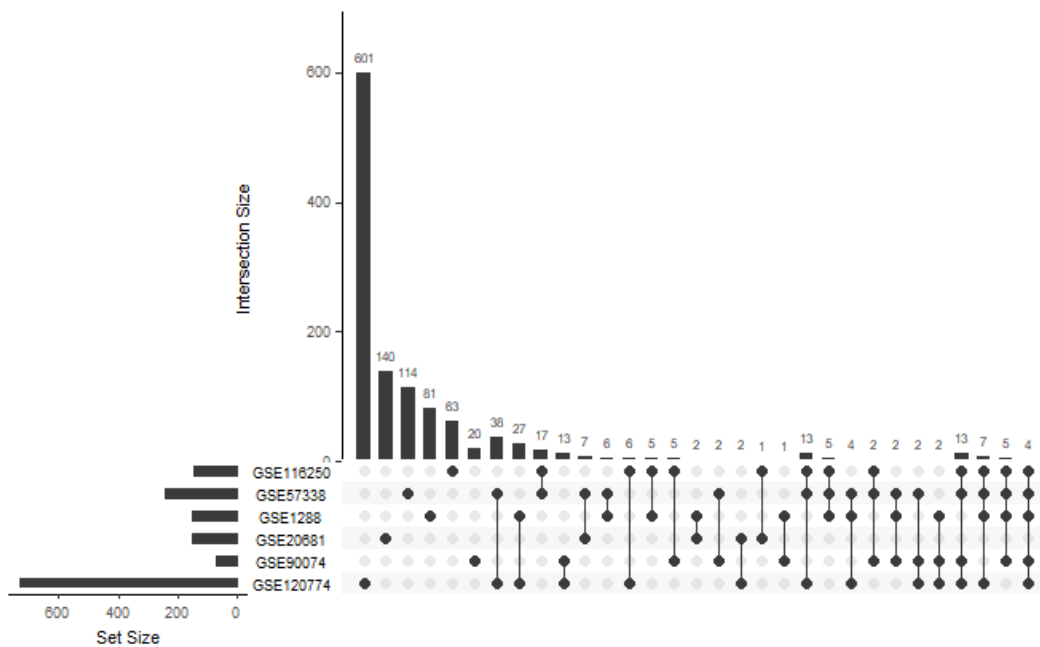
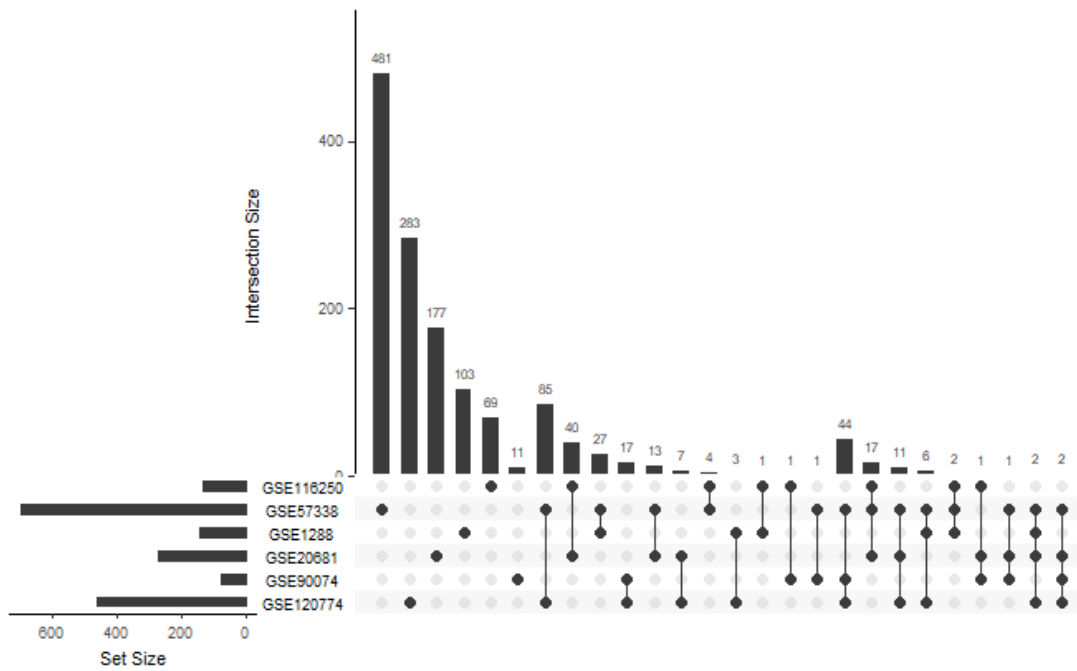


Figura 4.32: Componentes celulares descritos en la enfermedad de las arterias coronarias.



**Figura 4.33:** Funciones moleculares descritas para la enfermedad de las arterias coronarias.



**Figura 4.34:** Diagramas *Upset* para los procesos biológicos compartidos por los estudios de la enfermedad de las arterias coronarias. Arriba las funciones que tienen un LOR negativo y por tanto están sobrerrepresentadas en hombres con respecto a mujeres y abajo las funciones con un LOR negativo y sobrerrepresentadas en mujeres con respecto a los hombres.

El resto de diagramas *Upset* aparecen en el Anexo II (figuras a1 - a5).

#### 4.2.4. Análisis de rutas de señalización

Se realizó el análisis de rutas de señalización para 3 anotaciones, por un lado rutas KEGG, término GO y Uniprot. A continuación se muestran tablas resumen con el número de rutas significativas (p-valor ajustado por BH inferior a 0.05) para cada una de las enfermedades.

##### Estenosis aórtica

No hubo resultados significativos (p-valor ajustado por BH inferior a 0.05) para las rutas del estudio de la estenosis aórtica.

##### Enfermedad de las arterias coronarias

|           | GO | KEGG | UNIPROT |
|-----------|----|------|---------|
| GSE120774 | 0  | 0    | 0       |
| GSE90074  | 0  | 21   | 1       |
| GSE20681  | 0  | 0    | 0       |
| GSE1288   | 0  | 0    | 0       |
| GSE57338  | 0  | 1    | 0       |

**Tabla 4.20:** Resultados del análisis de rutas mediante Hipathia para los estudios de enfermedad de las arterias coronarias. Se muestran las rutas con un p-valor ajustado por BH inferior a 0.05.

A continuación se muestran las rutas significativas.

| Ruta                | logFC  | adj.P.Val   |
|---------------------|--------|-------------|
| Glycerol metabolism | 0.0499 | 0.00055     |
| P-hsa04914-23       | 0.022  | 0.04229     |
| P-hsa03320-31       | 0.0910 | 2.39035e-05 |
| P-hsa03320-40       | 0.0597 | 2.39035e-05 |
| P-hsa03320-9        | 0.0442 | 2.39035e-05 |
| P-hsa03320-29       | 0.0497 | 4.75567e-05 |

|               |         |             |
|---------------|---------|-------------|
| P-hsa03320-10 | 0.0497  | 4.75567e-05 |
| P-hsa04211-52 | 0.0530  | 5.42065e-05 |
| P-hsa03320-53 | 0.0494  | 5.55270e-05 |
| P-hsa04970-45 | -0.0259 | 0.00016     |
| P-hsa03320-60 | 0.0617  | 0.00016     |
| P-hsa04972-39 | -0.0025 | 0.00018     |
| P-hsa03320-38 | 0.0459  | 0.00021     |
| P-hsa04024-76 | 0.0119  | 0.00022     |
| P-hsa04740-39 | 0.0130  | 0.00043     |
| P-hsa03320-8  | 0.0618  | 0.00056     |
| P-hsa04072-57 | 0.0210  | 0.00488     |
| P-hsa03320-7  | 0.0412  | 0.00529     |
| P-hsa03320-39 | 0.0340  | 0.00572     |
| P-hsa04022-44 | 0.0130  | 0.00849     |
| P-hsa04022-49 | 0.0108  | 0.01999     |
| P-hsa05110-52 | 0.0058  | 0.03670     |
| P-hsa03320-59 | 0.0328  | 0.03900     |
| P-hsa03320-31 | 0.0910  | 2.39035     |

**Tabla 4.21:** Rutas significativas para la enfermedad de las arterias coronarias.

### 4.3. Metaanálisis

Por un lado se realizó un metaanálisis de genes con los resultados de expresión diferencial de los estudios de la enfermedad de las arterias coronarias, y por otro lado se realizó el metaanálisis de las funciones a partir de los resultados del análisis de GSEA para los diferentes términos GO por separado (BP, CC y MF), las rutas KEGG y los términos HPO asociados a fenotipos.

#### 4.3.1. Metaanálisis de genes

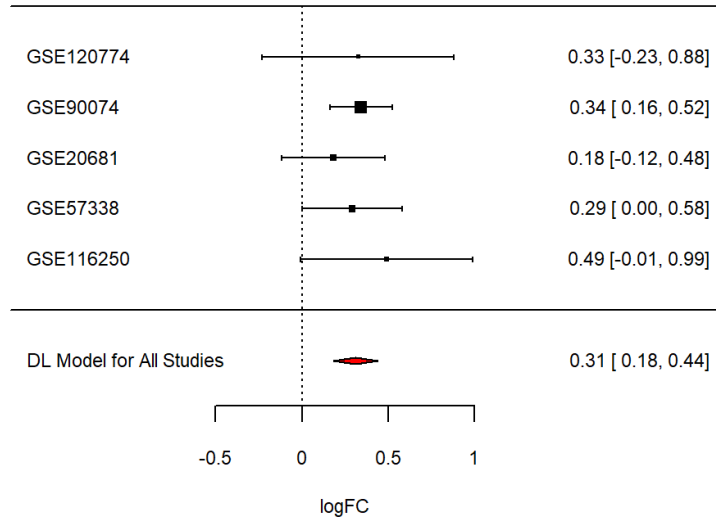
Para la enfermedad de las arterias coronarias se obtuvieron 3 genes con un p-valor ajustado por BH inferior a 0.05.

| ID     | Nombre                                      | LOR 95% I | LOR   | LOR 95% S | Adj P.val |
|--------|---|-----------|-------|-----------|-----------|
| 441178 | long intergenic non-protein coding RNA 2487 | 0.058     | 0.1   | 0.142     | 0.019     |
| 643008 | small integral membrane protein 5           | 0.253     | 0.421 | 0.59      | 0.017     |
| 83606  | guanylyl cyclase domain containing 1        | 0.185     | 0.313 | 0.441     | 0.017     |

**Tabla 4.22:** Genes con un p-valor ajustado por BH inferior a 0.05 para la enfermedad de las arterias coronarias. Se muestra el intervalo de confianza del LOR al 95% inferior (I), superior (S) y el p-valor ajustado.

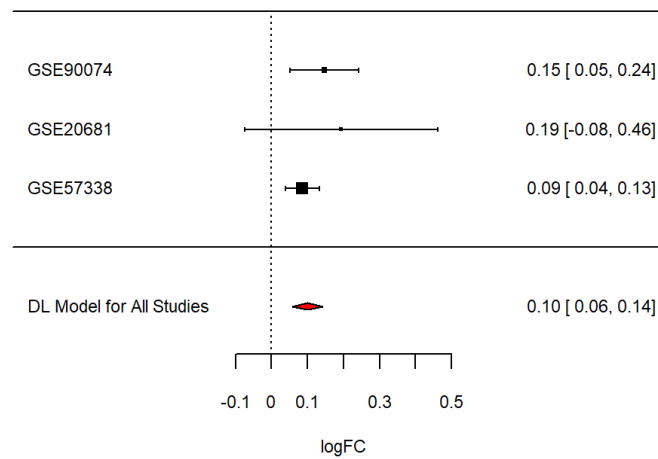
Los 3 genes presentan un valor de LOR positivo y por tanto, están más expresados en mujeres con la enfermedad que en hombres enfermos, cuando son comparados con sus respectivos grupos control. A continuación se muestran los diagramas de bosque para los genes significativos.

**83606**  
**guanylyl cyclase domain containing 1**



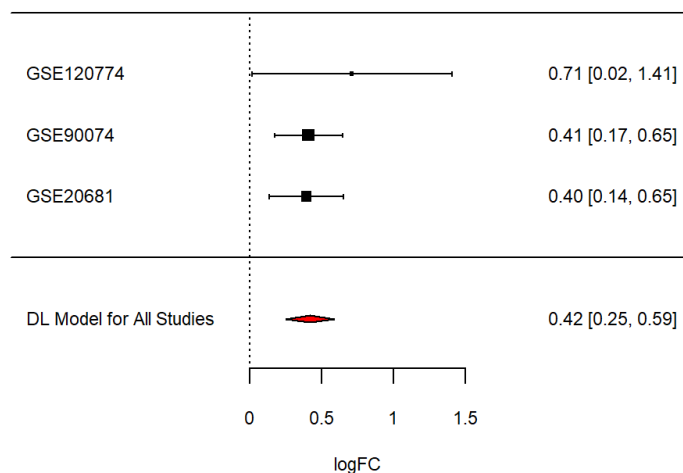
**Figura 4.35:** Diagrama de bosque para el gen con identificador Entrez 83606.

**441178**  
**long intergenic non-protein coding RNA 2487**



**Figura 4.36:** Diagrama de bosque para el gen con identificador Entrez 441178.

**643008**  
**small integral membrane protein 5**



**Figura 4.37:** Diagrama de bosque para el con ID de Entrez 644008.

En el Anexo II aparecen los gráficos de embudo con los resultados del MA (figuras a6 y a7 y a8).

### 4.3.2. Metaanálisis de funciones

A continuación se muestra una tabla resumen con los resultados del metaanálisis para cada una de las 3 ontologías de la GO, los términos KEGG y HPO que mostraron un p-valor ajustado por BH inferior a 0.05. Se agruparon en *Up* aquellas funciones con un LOR positivo y por tanto, una mayor actividad en mujeres con la enfermedad cardiovascular en relación a los hombres y *Down*, con mayor actividad en hombres.

|     | GO: BP |      | GO: CC |      | GO: MF |      | KEGG |      | HPO |      |
|-----|--------|------|--------|------|--------|------|------|------|-----|------|
|     | Up     | Down | Up     | Down | Up     | Down | Up   | Down | Up  | Down |
| CAD | 3      | 1    | 5      | 4    | 5      | 2    | 0    | 0    | 3   | 0    |

**Tabla 4.23:** Resumen de las funciones con un p-valor ajustado inferior a 0.05. CAD, enfermedad de las arterias coronarias.

## Procesos biológicos

A continuación se muestran las tablas con los procesos biológicos con un p-valor ajustado por BH inferior a 0.05.

| ID         | nombre   | interv_inf | Summary_L<br>OR | interv_su<br>p | p.ajust.fdr |
|------------|--|------------|-----------------|----------------|-------------|
| GO:0006091 | generation of precursor metabolites and energy | 0.1843     | 0.2973          | 0.4102         | 0.00079     |
| GO:0007005 | mitochondrion organization                     | 0.1556     | 0.2265          | 0.2975         | 2.4673e-06  |
| GO:0071305 | cellular response to vitamin D                 | -0.5924    | -0.4107         | -0.2290        | 0.01514     |
| GO:1990748 | cellular detoxification                        | 0.1251     | 0.2221          | 0.3191         | 0.01514     |

**Tabla 4.24:** Procesos biológicos significativos (p-valor ajustado inferior a 0.05) para la enfermedad de las arterias coronarias. `interv_inf` e `interv_sup` son el intervalo de confianza del 95% para el valor de LOR. `Summary_LOR` es el valor de LOR promedio y `p.ajust.fdr` es el p-valor ajustado.

## Componentes celulares

Seguidamente se muestran las tablas con los resultados obtenidos para los componentes celulares.

| ID         | nombre                          | interv_inf            | Summary_LOR | interv_sup            | p.ajust.fdr             |
|------------|---------------------------------|-----------------------|-------------|-----------------------|-------------------------|
| GO:0000786 | nucleosome                      | -0.2982               | -0.2062     | -0.1142               | 0.00217                 |
| GO:0000788 | nuclear nucleosome              | -0.4325               | -0.2859     | -0.1393               | 0.01304                 |
| GO:0005665 | RNA polymerase II, core complex | 0.3237331<br>88953925 | 0.6004      | 0.8772509798<br>10623 | 0.00333869<br>323965807 |
| GO:0005741 | mitochondrial outer membrane    | 0.0725                | 0.1617      | 0.2510                | 0.03337                 |
| GO:0019866 | organelle inner membrane        | 0.1903                | 0.2871      | 0.3839                | 4.86799e-06             |
| GO:0019867 | outer membrane                  | 0.0792                | 0.1510      | 0.2227                | 0.00486                 |
| GO:0031968 | organelle outer membrane        | 0.0765                | 0.1520      | 0.2275                | 0.00899                 |
| GO:0032993 | protein-DNA complex             | -0.2217               | -0.1596     | -0.0975               | 0.00018                 |
| GO:0044815 | DNA packaging complex           | -0.2920               | -0.2064     | -0.1208               | 0.00060                 |

**Tabla 4.25:** Componentes celulares significativos (p-valor ajustado inferior a 0.05) para la enfermedad de las arterias coronarias. *interv\_inf* e *interv\_sup* son el intervalo de confianza del 95% para el valor de LOR. *Summary\_LOR* es el valor de LOR promedio y *p.ajust.fdr* es el p-valor ajustado

## Funciones moleculares

A continuación se muestran los resultados para las funciones moleculares:

| ID         | nombre  | interv_inf | Summary_LOR | interv_sup | p.ajust.fdr |
|------------|---|------------|-------------|------------|-------------|
| GO:0008320 | protein transmembrane transporter activity              | 0.2458     | 0.4516      | 0.6573     | 0.01108     |
| GO:0008565 | protein transporter activity                            | 0.2302     | 0.4244      | 0.6185     | 0.01108     |
| GO:0016209 | antioxidant activity                                    | 0.1229     | 0.2398      | 0.3566     | 0.02328     |
| GO:0016684 | oxidoreductase activity, acting on peroxide as acceptor | 0.13137    | 0.2653      | 0.3992     | 0.02499     |
| GO:0022884 | macromolecule transmembrane transporter activity        | 0.1952     | 0.3889      | 0.5827     | 0.02499     |
| GO:0030676 | Rac guanyl-nucleotide exchange factor activity          | -0.5728    | -0.3775     | -0.1822    | 0.03035     |
| GO:0035004 | phosphatidylinositol 3-kinase activity                  | -0.7563    | -0.4967     | -0.2371    | 0.030365    |

**Tabla 4.26:** Funciones moleculares significativas (p-valor ajustado inferior a 0.05) para la enfermedad de las arterias coronarias. *interv\_inf* e *interv\_sup* son el intervalo de confianza del 95% para el valor de LOR. *Summary\_LOR* es el valor de LOR promedio y *p.ajust.fdr* es el p-valor ajustado

## Rutas Kegg

No se obtuvieron resultados significativos (p-valor ajustado por BH inferior a 0.05) para el metaanálisis de rutas KEGG para la enfermedad de las arterias coronarias.

## Fenotipos HPO

A continuación se muestran los resultados para los fenotipos de la HPO.

| ID         | nombre                                | interv_inf | Summary_LOR | interv_sup | p.ajust.fdr |
|------------|---------------------------------------|------------|-------------|------------|-------------|
| HP:0001081 | Cholelithiasis                        | 0.1371     | 0.2529      | 0.3687     | 0.02007     |
| HP:0002013 | Vomiting                              | 0.0988     | 0.1705      | 0.2423     | 0.00683     |
| HP:0012437 | Abnormal<br>gallbladder<br>morphology | 0.11044    | 0.2121      | 0.3138     | 0.03133     |

**Tabla 4.27:** Resultados significativos ( $p$ -valor ajustado por BH inferior a 0.05) para la enfermedad de las arterias coronarias para los fenotipos HPO. `interv_inf` e `interv_sup` son el intervalo de confianza del 95% para el valor de LOR. `Summary_LOR` es el valor de LOR promedio y `p.ajust.fdr` es el  $p$ -valor ajustado.

## 5. Discusión

### **Importancia de la inclusión de la perspectiva de sexo en los estudios biomédicos**

Las enfermedades cardiovasculares son una de las principales causas de muerte a nivel mundial. Estas patologías han sido clasificadas como enfermedades que afectan principalmente a hombres, por lo que las pruebas diagnósticas, marcadores de la enfermedad y tratamientos han sido enfocadas hacia este sexo<sup>3,4</sup>. Sin embargo, cada vez son más estudios los que demuestran que hay diferencias en la prevalencia, fisiopatología y tratamiento en las enfermedades cardiovasculares en cuanto al sexo y su consideración es un paso necesario e importante en el desarrollo y aplicación de la medicina personalizada.

Esta necesidad se evidencia cuando realizamos búsquedas y revisiones sistemáticas de estudios biomédicos cuyos datos están disponibles en los repositorios de acceso público. En muchos de estos estudios encontrados en relación a enfermedades cardiovasculares, no se presentaba la identificación del sexo en sus muestras o bien, el estudio estaba centrado en hombres, sin ninguna muestra representativa de mujeres. En definitiva, se constata una falta de inclusión de la perspectiva de sexo con una clara infrarrepresentación de mujeres en los estudios biomédicos en este grupo de patologías, incluso aunque el objetivo de la mayoría de estos trabajos fuera la detección de biomarcadores de la enfermedad en población general.

### **Principios FAIR**

Por este motivo, sumado a la necesaria estandarización de los datos en diferentes áreas de la ciencia, es por lo que diferentes expertos propusieron los principios FAIR<sup>45</sup> (*Findable, Accessible, Interoperable and Reusable*), para promover una buena práctica para la gestión y administración de datos científicos. La implantación de estos requerimientos son necesarios en cada vez más investigaciones, como por ejemplo las enmarcadas en el programa Horizonte 2020 de investigación e Innovación de la Unión Europea<sup>46</sup>. Su aplicación permitirá llevar a cabo estudios *in silico* libre de sesgos y que proporcionen una información relevante con una mayor potencia en los abordajes.

### **Caracterización de los mecanismos diferenciales por sexo de las enfermedades**

En este estudio, se ha procedido a analizar los datos disponibles en repositorios públicos, tanto a nivel individual como de forma conjunta con el metaanálisis con el fin de describir los procesos moleculares alterados de forma diferencial en hombres y mujeres en dos de

las enfermedades cardiovasculares más frecuentes y que mayor número de muertes producen a nivel mundial . Para ello, hemos tenido que trabajar con datos procedentes de diferentes proyectos, con diferentes objetivos, diferentes tipos de tejido, distintas plataformas y sin una nomenclatura estandarizada para los metadatos. Es por ello que la revisión sistemática y el metaanálisis han sido las herramientas de elección para integrar todos estos datos con el fin de obtener resultados más precisos y robustos<sup>23,26</sup>. Con ello, hemos podido encontrar un gran número de funciones alteradas de forma diferencial en la enfermedad de las arterias coronarias (tabla 4.23), sin embargo, algunas limitaciones han sido el número reducido de estudios seleccionados y el número de muestras, ya que en el caso de la estenosis aórtica, solo dos de los estudios cumplían los criterios de inclusión, por lo que finalmente no se ha realizado el metaanálisis en esta enfermedad (además en uno de los dos estudios, sólo se disponían de 1 mujer en el grupo control (tabla 4.3)).

### **Estenosis aórtica**

Como se ha comentado anteriormente, solo se ha realizado el análisis individual de uno de los dos estudios (E-MEXP-3684) por falta de muestras representativas de mujer control en el estudio GSE10161.

En el análisis individual del estudio seleccionado se encontraron un gran número de funciones diferencialmente representadas en hombres y en mujeres (tabla 4.15). Entre los procesos biológicos con resultado significativo, en mujeres hay una mayor representación de las funciones relacionadas con el metabolismo mitocondrial (figura 4.29, tabla a3), al igual que aparecen componentes celulares relacionados con la cadena transportadora de electrones en los resultados significativos (figura 4.30, tabla a4), todos ellos con un valor LOR positivo. Los resultados de las funciones moleculares con resultado significativo (tabla a5) también muestran funciones relacionadas con el metabolismo energético por parte de las mitocondrias. Todos estos resultados apoyan a que en mujeres hay una mayor preservación de la función mitocondrial con respecto a los hombres.

Hay estudios que confirman que en mujeres con enfermedades cardiovasculares hay una mejor preservación de la función mitocondrial en relación a los hombres<sup>47</sup>, por lo que apoyaría los resultados obtenidos en este estudio. En concreto, el metabolismo mitocondrial y la producción de energía también están regulados por los estrógenos, que actúan sobre los receptores de estrógenos presentes en las mitocondrias<sup>48</sup>, llevando a cabo un aumento de la expresión de proteínas mitocondriales y la regulación de su función. Por este motivo, también se ha reportado en un estudio que el tratamiento con estrógenos puede recuperar la función mitocondrial cardíaca de mujeres con menopausia<sup>49</sup>.

En hombres destacan los procesos biológicos relacionados con la regulación del sistema inmune, principalmente la quimiotaxis de linfocitos T (GO:0010818) y su regulación negativa

(GO:0045060), la secreción de citoquinas (GO:0050707) y la respuesta inmune mediada por inmunoglobulinas (GO:0016064), que tuvieron los valores de LOR más negativos (tabla a3). En componentes celulares también aparecen componentes relacionados con el sistema inmune, como son los receptores de linfocitos T (GO:0042101) y el complejo mayor de histocompatibilidad (GO:0042613) y funciones moleculares relacionadas con el reconocimiento de patrones celulares mediadas por receptor (GO:0038187). Se ha reportado infiltración de linfocitos T en pacientes con estenosis aórtica, que conlleva un aumento del daño producido por la enfermedad<sup>50</sup>.

En un estudio bioinformático reciente que sigue una estrategia similar al presente estudio, en el que se analizan los marcadores genéticos de la estenosis aórtica también aparecen la respuesta inmune y la quimiotaxis como principales funciones alteradas por la enfermedad<sup>51</sup>, por lo que apoyaría los resultados obtenidos en los análisis realizados en este estudio, sin embargo, no se aplica la perspectiva de sexo. También hay estudios que demuestran que hay diferencias en cuanto al sexo en la expresión de genes reguladores de la inflamación y la fibrosis, con una menor activación de estos procesos en mujeres con respecto a los hombres<sup>52,53</sup>.

### **Enfermedad de las arterias coronarias**

En el caso de los estudios de la enfermedad de las arterias coronarias, no se han encontrado funciones compartidas por todos los estudios (figuras 4.34 y figuras a1 - a5) para ninguna de las diferentes anotaciones funcionales utilizadas. Sin embargo, el MA sí que mostró algunas funciones significativas (p-valor ajustado inferior a 0.05), demostrando que es una potente herramienta para analizar datos de diferentes estudios.

Entre estas funciones, destacan los procesos biológicos relacionados con la organización de mitocondrias (GO:0007005), generación de metabolitos precursores y energía (GO:0006091) y procesos de detoxificación celular (GO:1990748). Todos ellos están ligeramente sobrerrepresentados en mujeres con respecto a hombres (LOR positivo). En hombres, solo se ha encontrado el proceso metabólico relacionado con la respuesta celular a la vitamina D (GO:0071305) como función más sobrerrepresentada en hombres con respecto a mujeres. Diferentes estudios<sup>54,55</sup> demuestran los efectos beneficiosos de la vitamina D contra la enfermedad de las arterias coronarias. La deficiencia de vitamina D está asociada a varias enfermedades inflamatorias crónicas como son la diabetes y la obesidad, que son fuertes factores de riesgo para desarrollar la enfermedad de las arterias coronarias<sup>55</sup>. Hay estudios que demuestran que en mujeres con enfermedades cardiovasculares hay una mejor preservación de la función mitocondrial y un mayor número de procesos de detoxificación<sup>47</sup>, por lo que se generan menos radicales libres que puedan

afectar al desarrollo de la enfermedad y estos resultados apoyan los obtenidos en el MA.

Los resultados para los componentes celulares (tabla 4.25) apuntan a que hay una mayor representación de funciones relacionadas con las membranas internas y externas de los orgánulos, entre ellos las mitocondrias (GO:0005741) en mujeres en relación con hombres, en los que en hombres los componentes celulares significativos están más relacionados con el nucleosoma (GO:0000786, GO:0000788), que es el orgánulo encargado de la organización de la cromatina. También hay una mayor representación de funciones de interacción ADN-proteína (GO:0032993) y su empaquetamiento (GO:0044815) en hombres con respecto a mujeres. Hay algunos estudios<sup>47</sup> que demuestran cómo cambios epigenéticos pueden intervenir en la enfermedad de las arterias coronarias, aunque no son numerosos.

Entre las funciones moleculares alteradas en la enfermedad de las arterias coronarias, los resultados apuntan a que en mujeres hay una mayor representación de funciones relacionadas con el proceso de transporte transmembrana (GO:0008320 y GO:0008565), funciones antioxidantes (GO:0016209) y actividad oxidorreductasa (GO:0016684), por lo que podría estar relacionado con una mejor preservación de estas funciones como se ha comentado anteriormente. Hay estudios<sup>47,56</sup> que demuestran cómo en pacientes con la enfermedad de las arterias coronarias hay un aumento de los fenómenos de peroxidación de lípidos de membrana y enfermedades como la Diabetes Mellitus, hipertensión y hábitos como fumar aumenta la producción de radicales libres y estrés oxidativo, con lo que inducen el daño en la vasculatura de las arterias<sup>47</sup>.

En hombres en cambio, los resultados del MA para las funciones moleculares apuntan a funciones de transducción de señales, con la actividad de Rac Guanilil ciclasa (GO:0030676) y fosfatidil inositol trifosfato kinasa (GO:0035004).

Por lo tanto, como se ha comentado anteriormente, los resultados del MA muestran como hay un gran número de funciones que están alteradas de forma diferencial en hombres y en mujeres en diferentes enfermedades cardiovasculares. Dada la importancia de estas enfermedades en la sociedad, siendo la primera causa de muerte en los países de ingreso medio<sup>57</sup>, es importante la inclusión del sexo en los estudios biomédicos, ya que se ha encontrado un vacío representativo en la información del sexo en los estudios, lo que puede dificultar la generalización y aplicación de los avances científicos a la población general y además, conocer las diferencias en cuanto al sexo podría permitir un mejor diagnóstico y desarrollo de tratamientos específicos.

## 6. Conclusiones

- 1) El metaanálisis es una herramienta robusta para la evaluación e integración de datos ómicos procedente de la selección sistemática de estudios disponibles en bases de datos públicos.
- 2) Los resultados del análisis del estudio individual de estenosis aórtica apuntan a que hay una sobrerrepresentación de funciones relacionadas con el sistema inmunitario en hombres, mientras que en mujeres hay un mayor número de funciones relacionadas con el mantenimiento del metabolismo mitocondrial.
- 3) El MA para la enfermedad de las arterias coronarias identifica una mayor sobrerrepresentación de funciones relacionadas con la organización y función mitocondrial en mujeres, además de procesos relacionados con la detoxificación, mientras que en hombres, las funciones están más relacionadas con la estructura y cambios epigenéticos en el ADN y la transducción de señales.
- 4) La estrategia seguida en este estudio ha permitido detectar y caracterizar molecularmente las diferencias en cuanto al sexo en pacientes con enfermedades cardiovasculares mediante un abordaje *in silico* de datos ómicos.
- 5) Es necesaria la inclusión del sexo en los estudios biomédicos con el fin de obtener resultados más precisos para el desarrollo y la aplicación de la Medicina Personalizada.
- 6) Este trabajo es el punto de partida para la revisión actualizada de otros estudios de estenosis aórtica que se incorporen próximamente en los repositorios públicos, y que permitan un abordaje completo como el realizado en la enfermedad de las arterias coronarias. Asimismo también constituye una primera parte de un proyecto donde se evaluarán los mecanismos diferenciales por sexo en otras enfermedades cardiovasculares de interés.



## Bibliografía:

1. García M. Factores de riesgo cardiovascular desde la perspectiva de sexo y género. *Rev Colomb Cardiol*. 2018;25:8-12. doi:10.1016/j.rccar.2017.11.021
2. Oertelt-Prigione S, Prescott E, Franconi F, et al. Gender in cardiovascular diseases: impact on clinical manifestations, management, and outcomes. :16.
3. Humphries KH, Izadnegahdar M, Sedlak T, et al. Sex differences in cardiovascular disease – Impact on care and outcomes. *Front Neuroendocrinol*. 2017;46:46-70. doi:10.1016/j.yfrne.2017.04.001
4. Ruiz Cantero MT. *Perspectiva de género en medicina.*; 2019.
5. Aortic Stenosis Overview | American Heart Association. Accessed September 6, 2020. <https://www.heart.org/en/health-topics/heart-valve-problems-and-disease/heart-valve-problems-and-causes/problem-aortic-valve-stenosis>
6. Administrator. Valvulopatía aórtica. Fundación Española del Corazón. Accessed September 6, 2020. <https://fundaciondelcorazon.com/informacion-para-pacientes/enfermedades-cardiovasculares/valvulopatias/valvulopatias-aortica.html>
7. Bhandari B, Masood W. Ischemic Cardiomyopathy. In: *StatPearls*. StatPearls Publishing; 2020. Accessed September 6, 2020. <http://www.ncbi.nlm.nih.gov/books/NBK537301/>
8. Coronary Heart Disease | NHLBI, NIH. Accessed September 6, 2020. <https://www.nhlbi.nih.gov/health-topics/coronary-heart-disease>
9. Divoky L, Maran A, Ramu B. Gender Differences in Ischemic Cardiomyopathy. *Curr Atheroscler Rep*. 2018;20(10):50. doi:10.1007/s11883-018-0750-x
10. Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P. *Molecular Biology of the Cell*. 4th ed. Garland Science; 2002.
11. Fisher M. Lehninger Principles of Biochemistry, 3rd edition; By David L. Nelson and Michael M. Cox. *Chem Educ*. 2001;6(1):69-70. doi:10.1007/s00897000455a
12. Meta-analytic principal component analysis in integrative omics application | Bioinformatics | Oxford Academic. Accessed September 6, 2020. <https://academic.oup.com/bioinformatics/article/34/8/1321/4653703>
13. Schena M. Genome analysis with gene expression microarrays. *BioEssays News Rev Mol Cell Dev Biol*. 1996;18(5):427-431. doi:10.1002/bies.950180513
14. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet*. 2009;10(1):57-63. doi:10.1038/nrg2484
15. García FG. Métodos de análisis de enriquecimiento funcional en estudios genómicos. :191.
16. Ashburner M, Ball CA, Blake JA, et al. Gene Ontology: tool for the unification of biology. *Nat Genet*. 2000;25(1):25-29. doi:10.1038/75556
17. Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res*. 2000;28(1):27-30.
18. Fabregat A, Jupe S, Matthews L, et al. The Reactome Pathway Knowledgebase. *Nucleic Acids Res*. 2018;46(Database issue):D649-D655. doi:10.1093/nar/gkx1132
19. Köhler S, Vasilevsky NA, Engelstad M, et al. The Human Phenotype Ontology in 2017. *Nucleic Acids Res*. 2017;45(Database issue):D865-D876. doi:10.1093/nar/gkw1039
20. Maleki F, Ovens K, Hogan DJ, Kusalik AJ. Gene Set Analysis: Challenges, Opportunities, and Future Research. *Front Genet*. 2020;11. doi:10.3389/fgene.2020.00654
21. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*. 2005;102(43):15545-15550. doi:10.1073/pnas.0506580102

22. Hidalgo MR, Cubuk C, Amadoz A, Salavert F, Carbonell-Caballero J, Dopazo J. High throughput estimation of functional cell activities reveals disease mechanisms and predicts relevant clinical outcomes. *Oncotarget*. 2017;8(3):5160-5178. doi:10.18632/oncotarget.14107
23. Normand SL. Meta-analysis: formulating, evaluating, combining, and reporting. *Stat Med*. 1999;18(3):321-359. doi:10.1002/(sici)1097-0258(19990215)18:3<321::aid-sim28>3.0.co;2-p
24. Panagiotou OA, Willer CJ, Hirschhorn JN, Ioannidis JPA. The Power of Meta-Analysis in Genome Wide Association Studies. *Annu Rev Genomics Hum Genet*. 2013;14:441-465. doi:10.1146/annurev-genom-091212-153520
25. Díaz RB, Cahua MC. Introducción al meta-análisis tradicional. :7.
26. Haidich AB. Meta-analysis in medical research. *Hippokratia*. 2010;14(Suppl 1):29-37.
27. Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *J Clin Epidemiol*. 2009;62(10):e1-34. doi:10.1016/j.jclinepi.2009.06.006
28. Moher D, Shamseer L, Clarke M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Syst Rev*. 2015;4:1. doi:10.1186/2046-4053-4-1
29. Brazma A, Kapushesky M, Parkinson H, Sarkans U, Shojatalab M. Data storage and analysis in ArrayExpress. *Methods Enzymol*. 2006;411:370-386. doi:10.1016/S0076-6879(06)11020-4
30. Davis S, Meltzer PS. GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor. *Bioinforma Oxf Engl*. 2007;23(14):1846-1847. doi:10.1093/bioinformatics/btm254
31. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J R Stat Soc Ser B Methodol*. 1995;57(1):289-300.
32. Montaner D, Dopazo J. Multidimensional gene set analysis of genomic data. *PloS One*. 2010;5(4):e10348. doi:10.1371/journal.pone.0010348
33. Peña-Chilet M, Esteban-Medina M, Falco MM, et al. Using mechanistic models for the clinical interpretation of complex genomic variation. *Sci Rep*. 2019;9(1):18937. doi:10.1038/s41598-019-55454-7
34. Viechtbauer W. Conducting Meta-Analyses in R with the **metafor** Package. *J Stat Softw*. 2010;36(3). doi:10.18637/jss.v036.i03
35. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials*. 1986;7(3):177-188. doi:10.1016/0197-2456(86)90046-2
36. Hedges LV, Gurevitch J, Curtis PS. The Meta-Analysis of Response Ratios in Experimental Ecology. *Ecology*. 1999;80(4):1150-1156. doi:10.2307/177062
37. *Methods of Meta-Analysis: Correcting Error and Bias in Research Findings - SAGE Research Methods*. Accessed September 6, 2020. <http://methods.sagepub.com/book/methods-of-meta-analysis-3e>
38. Integrated genomic approaches implicate osteoglycin (Ogn) in the regulation of left ventricular mass - PubMed. Accessed September 6, 2020. <https://pubmed.ncbi.nlm.nih.gov/18443592/>
39. Fitzgibbons TP, Lee N, Tran K-V, et al. Coronary disease is not associated with robust alterations in inflammatory gene expression in human epicardial fat. *JCI Insight*. 2019;4(20). doi:10.1172/jci.insight.124859
40. Ravi S, Schuck RN, Hilliard E, et al. Clinical Evidence Supports a Protective Role for CXCL5 in Coronary Artery Disease. *Am J Pathol*. 2017;187(12):2895-2911. doi:10.1016/j.ajpath.2017.08.006
41. Elashoff MR, Wingrove JA, Beineke P, et al. Development of a blood-based gene expression algorithm for assessment of obstructive coronary artery disease in non-diabetic patients. *BMC Med Genomics*. 2011;4:26. doi:10.1186/1755-8794-4-26

42. Sinnaeve PR, Donahue MP, Grass P, et al. Gene expression patterns in peripheral blood correlate with the extent of coronary artery disease. *PloS One*. 2009;4(9):e7037. doi:10.1371/journal.pone.0007037
43. Liu Y, Morley M, Brandimarto J, et al. RNA-Seq identifies novel myocardial gene expression signatures of heart failure. *Genomics*. 2015;105(2):83-89. doi:10.1016/j.ygeno.2014.12.002
44. Sweet ME, Cocciolo A, Slavov D, et al. Transcriptome analysis of human heart failure reveals dysregulated cell adhesion in dilated cardiomyopathy and activated immune pathways in ischemic heart failure. *BMC Genomics*. 2018;19(1):812. doi:10.1186/s12864-018-5213-9
45. Wilkinson MD, Dumontier M, Aalbersberg IJ, et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data*. 2016;3. doi:10.1038/sdata.2016.18
46. Open access & Data management - H2020 Online Manual. Accessed September 6, 2020. [https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-dissemination\\_en.htm](https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-dissemination_en.htm)
47. Regitz-Zagrosek V, Kararigas G. Mechanistic Pathways of Sex Differences in Cardiovascular Disease. *Physiol Rev*. 2017;97(1):1-37. doi:10.1152/physrev.00021.2015
48. Mitochondrial localization of ERalpha and ERbeta in human MCF7 cells - PubMed. Accessed September 6, 2020. <https://pubmed.ncbi.nlm.nih.gov/14736707/>
49. Rattanasopa C, Phungphong S, Wattanapermpool J, Bupha-Intr T. Significant role of estrogen in maintaining cardiac mitochondrial functions. *J Steroid Biochem Mol Biol*. 2015;147:1-9. doi:10.1016/j.jsbmb.2014.11.009
50. Wu HD, Maurer MS, Friedman RA, et al. The lymphocytic infiltration in calcific aortic stenosis predominantly consists of clonally expanded T cells. *J Immunol Baltim Md 1950*. 2007;178(8):5329-5339. doi:10.4049/jimmunol.178.8.5329
51. Teng P, Xu X, Ni C, et al. Identification of key genes in calcific aortic valve disease by integrated bioinformatics analysis. *Medicine (Baltimore)*. 2020;99(29):e21286. doi:10.1097/MD.00000000000021286
52. Gaignebet L, Kańduła MM, Lehmann D, Knosalla C, Kreil DP, Kararigas G. Sex-Specific Human Cardiomyocyte Gene Regulation in Left Ventricular Pressure Overload. *Mayo Clin Proc*. 2020;95(4):688-697. doi:10.1016/j.mayocp.2019.11.026
53. Kararigas G, Dworatzek E, Petrov G, et al. Sex-dependent regulation of fibrosis and inflammation in human left ventricular remodelling under pressure overload. *Eur J Heart Fail*. 2014;16(11):1160-1167. doi:10.1002/ejhf.171
54. Legarth C, Grimm D, Krüger M, Infanger M, Wehland M. Potential Beneficial Effects of Vitamin D in Coronary Artery Disease. *Nutrients*. 2019;12(1). doi:10.3390/nu12010099
55. Rai V, Agrawal DK. Role of Vitamin D in Cardiovascular Diseases. *Endocrinol Metab Clin North Am*. 2017;46(4):1039-1059. doi:10.1016/j.ecl.2017.07.009
56. Namazi G, Jamshidi Rad S, Attar AM, et al. Increased membrane lipid peroxidation and decreased Na<sup>+</sup>/K<sup>+</sup>-ATPase activity in erythrocytes of patients with stable coronary artery disease. *Coron Artery Dis*. 2015;26(3):239-244. doi:10.1097/MCA.0000000000000196
57. Las 10 principales causas de defunción. Accessed September 6, 2020. <https://www.who.int/es/news-room/fact-sheets/detail/the-top-10-causes-of-death>



## Anexo I: Tablas

| Software | Versión |
|----------|---------|
| R        | 3.5.1   |

**Tabla a1:** Software utilizado durante el trabajo.

| Paquete                       | Versión |
|-------------------------------|---------|
| Biobase                       | 2.42.0  |
| AnnotationDbi                 | 1.44.0  |
| ArrayExpress                  | 1.42.0  |
| biomaRt                       | 2.38.0  |
| clusterProfiler               | 3.10.1  |
| dplyr                         | 0.8.3   |
| edgeR                         | 3.24.3  |
| GEOquery                      | 2.50.5  |
| ggdendro                      | 0.1-20  |
| ggplot2                       | 3.1.0   |
| GO.db                         | 3.7.0   |
| hgu133a.db                    | 3.2.3   |
| hgu133a2.db                   | 3.2.3   |
| hgug4112a.db                  | 3.2.3   |
| hgug4110b.db                  | 3.2.3   |
| hipathia                      | 1.4.1   |
| hugene10stranscriptcluster.db | 8.7.0   |
| KEGG.db                       | 3.2.3   |
| limma                         | 3.36.5  |
| mdgsa                         | 1.14.0  |
| metafor                       | 2.4-0   |
| org.Hs.eg.db                  | 3.7.0   |
| SummarizedExperiment          | 1.12.0  |
| UpSetR                        | 1.4.0   |
| venn                          | 1.9     |

**Tabla a2:** Paquetes utilizados durante el desarrollo del trabajo.

| Procesos biológicos |   |         |                       |
|---------------------|---|---------|-----------------------|
| ID                  | nombre  | lor     | padj                  |
| GO:0045060          | negative thymic T cell selection                            | -1.1828 | 0.000750011520474433  |
| GO:0045577          | regulation of B cell differentiation                        | -1.1454 | 0.0027710480511049    |
| GO:0050707          | regulation of cytokine secretion                            | -1.1203 | 0.00397884858864327   |
| GO:0010818          | T cell chemotaxis   | -1.0847 | 0.00156586432952338   |
| GO:0016064          | immunoglobulin mediated immune response                     | -1.0780 | 0.000271293061527686  |
| GO:0006123          | mitochondrial electron transport, cytochrome c to oxygen    | 1.8334  | 2.20778267062629e-06  |
| GO:0015986          | ATP synthesis coupled proton transport                      | 1.8974  | 1.32125697617312e-14  |
| GO:0032981          | mitochondrial respiratory chain complex I assembly          | 1.9176  | 9.69012377119649e-103 |
| GO:0006120          | mitochondrial electron transport, NADH to ubiquinone        | 1.9933  | 2.3090880546585e-93   |
| GO:0006122          | mitochondrial electron transport, ubiquinol to cytochrome c | 2.0948  | 9.54620171095298e-34  |

**Tabla a3:** Selección de procesos biológicos significativos para la estenosis aórtica.

| Componentes Celulares |  |         |                       |
|-----------------------|--|---------|-----------------------|
| ID                    | nombre   | lor     | padj                  |
| GO:0042101            | T cell receptor complex  | -1.3089 | 8.75292224312854e-05  |
| GO:0043020            | NADPH oxidase complex  | -1.0209 | 0.000700918529222509  |
| GO:0008305            | integrin complex   | -0.8708 | 4.27970766621263e-05  |
| GO:0045095            | keratin filament   | -0.8273 | 1.90027860063946e-12  |
| GO:0042613            | MHC class II protein complex   | -0.8221 | 0.00477741853452385   |
| GO:0005750            | mitochondrial respiratory chain complex III                                  | 1.9671  | 1.09678810650007e-25  |
| GO:0045263            | proton-transporting ATP synthase complex, coupling factor F(o)               | 1.9725  | 1.00082224628539e-09  |
| GO:0070469            | respiratory chain  | 2.0017  | 4.99464928171677e-104 |
| GO:0005753            | mitochondrial proton-transporting ATP synthase complex                       | 2.0239  | 1.15293171403073e-19  |
| GO:0000276            | mitochondrial proton-transporting ATP synthase complex, coupling factor F(o) | 2.2056  | 7.956984569804e-61    |

**Tabla a4:** Selección de componentes celulares significativos para la estenosis aórtica.

| Funciones Moleculares |   |                    |                      |
|-----------------------|---|--------------------|----------------------|
| ID                    | nombre  | lor                | padj                 |
| GO:0071723            | lipopeptide binding   | -1.26656357833959  | 6.89518962890886e-05 |
| GO:0038187            | pattern recognition receptor activity                           | -1.14030137428015  | 0.000400108676570827 |
| GO:0008381            | mechanosensitive ion channel activity                           | -1.09054343879804  | 0.00288539727111198  |
| GO:0038024            | cargo receptor activity   | -1.03717758196001  | 0.00123306832121558  |
| GO:0016175            | superoxide-generating NADPH oxidase activity                    | -0.994108067701126 | 0.00872812284666778  |
| GO:0003735            | structural constituent of ribosome                              | 1.32245382403338   | 4.05335089750574e-66 |
| GO:0008276            | protein methyltransferase activity                              | 1.32891065098414   | 1.25541730350868e-07 |
| GO:0046933            | proton-transporting ATP synthase activity, rotational mechanism | 1.66193772939245   | 4.51861162088849e-08 |
| GO:0003954            | NADH dehydrogenase activity                                     | 1.69238105854673   | 7.53416168036653e-24 |
| GO:0008137            | NADH dehydrogenase (ubiquinone) activity                        | 1.91123073674853   | 3.80468694222503e-19 |

**Tabla a5:** Selección de funciones moleculares significativas para la estenosis aórtica.

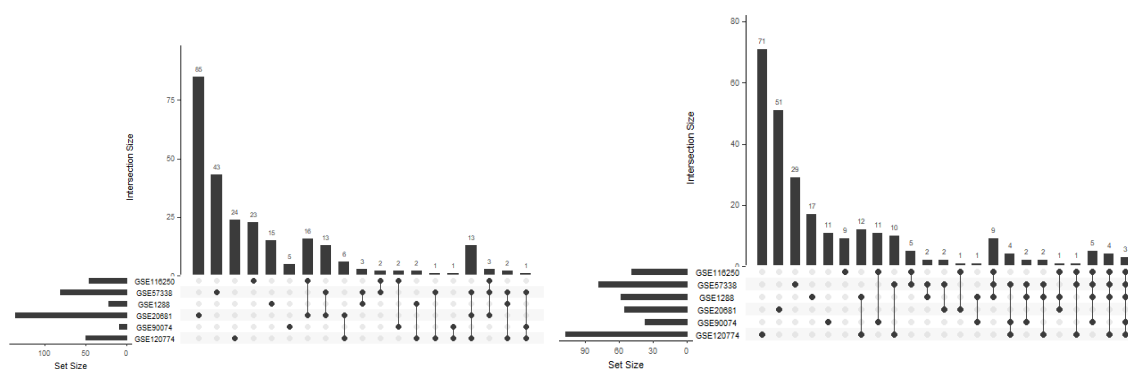
| Rutas KEGG |                                 |                    |                      |
|------------|---------------------------------|--------------------|----------------------|
| ID         | nombre                          | lor                | padj                 |
| 5340       | Primary immunodeficiency        | -0.854490500963436 | 8.07299627185268e-06 |
| 5332       | Graft-versus-host disease       | -0.795072039334505 | 3.90794616197172e-05 |
| 5330       | Allograft rejection             | -0.767311838356518 | 9.51708511394793e-05 |
| 5150       | Staphylococcus aureus infection | -0.765828193510989 | 6.22628845287202e-07 |
| 920        | Sulfur metabolism               | -0.714728703884114 | 0.0482308300114217   |
| 4260       | Cardiac muscle contraction      | 0.993684793446084  | 3.71467516451591e-11 |
| 5016       | Huntington disease              | 1.01854302902146   | 2.61107088205264e-29 |
| 3010       | Ribosome                        | 1.29919143893833   | 2.5175116143707e-36  |
| 5012       | Parkinson disease               | 1.41807791120547   | 4.80622353124916e-20 |
| 190        | Oxidative phosphorylation       | 1.4259357893096    | 2.57430168094056e-15 |

**Tabla a6:** Rutas KEGG significativas para la estenosis aórtica.

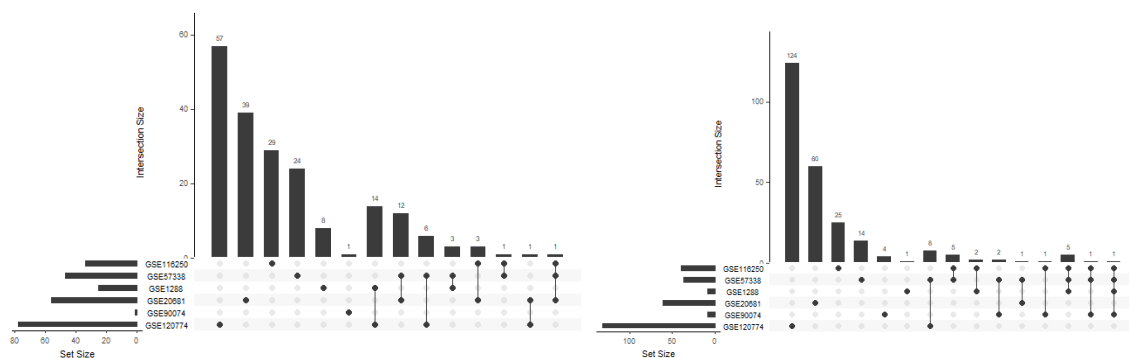
| Reactome      |                   |                       |
|---------------|-------------------|-----------------------|
| ID            | lor               | padj                  |
| R-HSA-2029481 | -1.22788698506382 | 0.000993834346450937  |
| R-HSA-202430  | -1.19220273270336 | 7.36031356841407e-08  |
| R-HSA-202427  | -1.17148361076996 | 7.39514874964144e-09  |
| R-HSA-2453864 | -1.00493439546486 | 0.000284853035143232  |
| R-HSA-2474795 | -1.00493439546486 | 0.000284853035143232  |
| R-HSA-1428517 | 1.54921792747746  | 6.47776070582661e-74  |
| R-HSA-163210  | 1.82481314525744  | 2.43803211426429e-12  |
| R-HSA-163200  | 1.830365991358    | 5.04808332266514e-84  |
| R-HSA-611105  | 1.91014253097134  | 1.49861918593415e-107 |
| R-HSA-6799198 | 1.97680907869365  | 6.50720183917935e-124 |

**Tabla a7:** Rutas Reactome significativas para la estenosis aórtica.

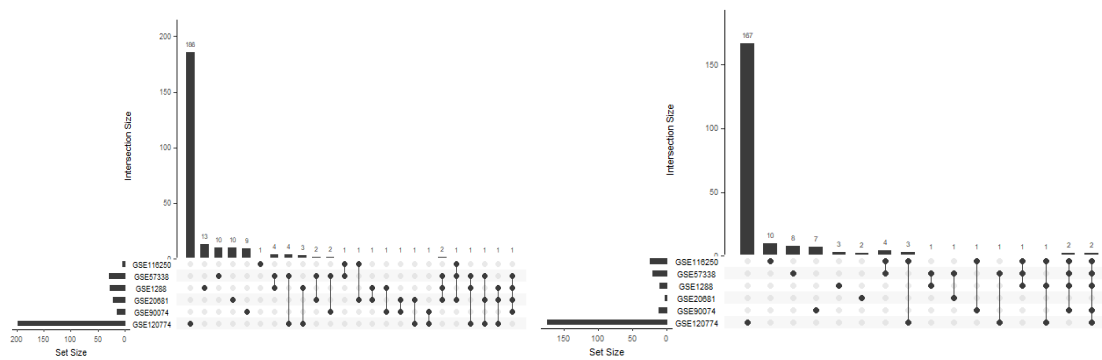
# Anexo II: Figuras



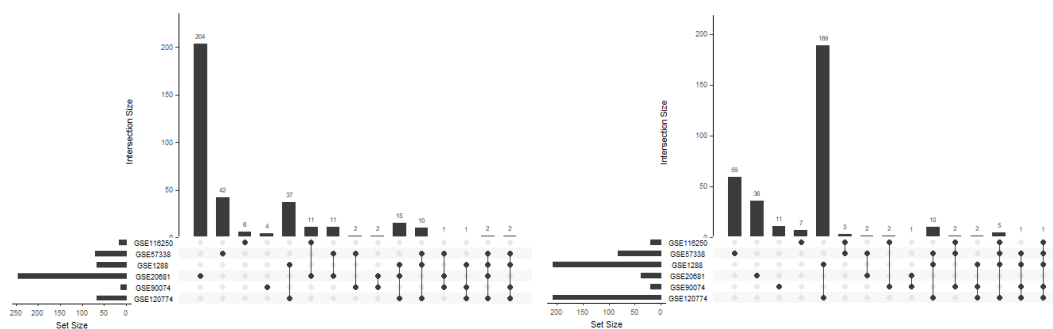
**Figura a1:** Diagramas *Upset* para los componentes celulares compartidos por los estudios de enfermedad de las arterias coronarias. A la izquierda las funciones con LOR negativo y a la derecha las funciones con LOR positivo.



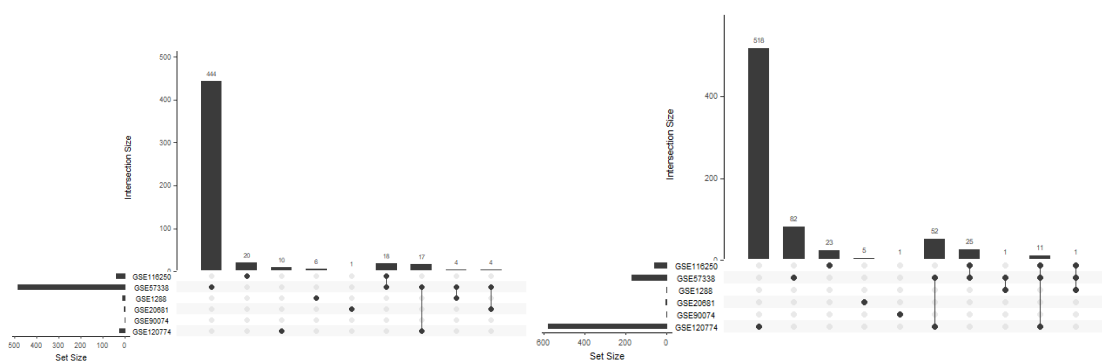
**Figura a2:** Diagramas *Upset* para las funciones moleculares compartidas por los estudios de enfermedad de las arterias coronarias. A la izquierda las funciones con un LOR negativo y a la derecha las funciones con un LOR positivo.



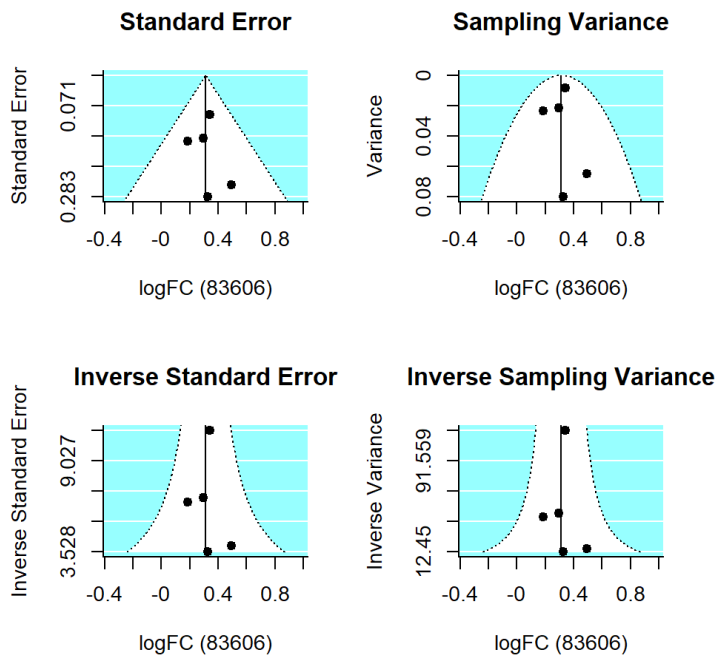
**Figura a3:** Diagramas *Upset* para las rutas KEGG compartidas por los estudios de enfermedad de las arterias coronarias. A la izquierda las rutas con un LOR negativo y a la derecha las rutas con un LOR positivo.



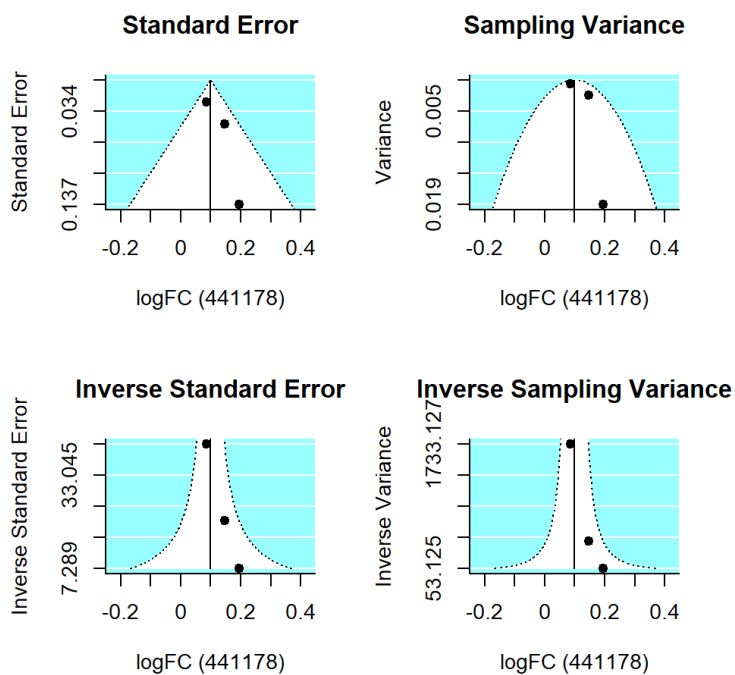
**Figura a4:** Diagramas *Upset* para las rutas Reactome compartidas por los diferentes estudios de enfermedad de las arterias coronarias. A la izquierda las funciones con LOR negativo y a la derecha las funciones con un LOR positivo.



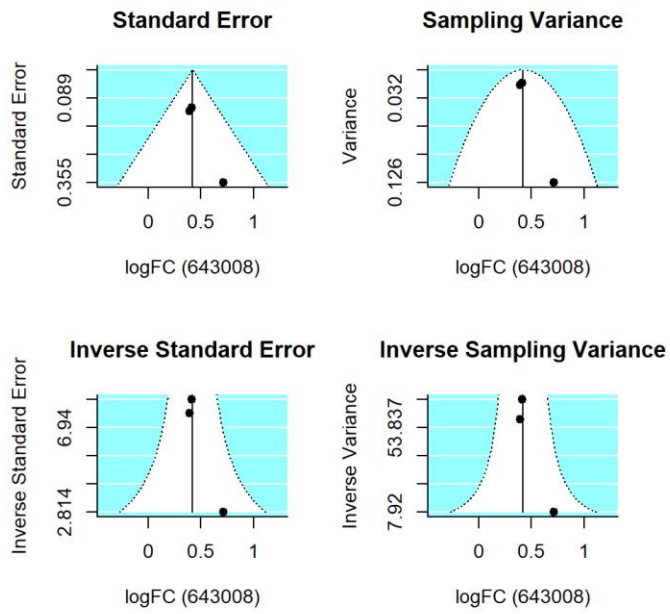
**Figura a5:** Diagramas *Upset* para las funciones HPO para la enfermedad de las arterias coronarias. A la izquierda las funciones con un LOR negativo y a la derecha las funciones con un LOR positivo.



**Figura 6:** Resultados del MA para el gen con identificador Entrez 83606.



**Figura a7:** Resultados del MA para el gen con identificador Entrez 441178.



**Figura a7:** Resultados del MA para el gen con el identificador Entrez 643006.